REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggesstions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any oenalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
12-01-2018	Final Report	17-Jan-2017 - 16-Jan-2018

4. TITLE AND SUBTITLE
Final Report: Decoding the Principles of Emergence and Resiliency in Biological Collective Systems - A Multi-scale Approach

5a. CONTRACT NUMBER
W911NF-17-1-0076

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of Southern California
Contracts & Grants
3720 S. Flower St.
Los Angeles, CA 90089 -0701

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
70347-EG-DRP.4

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Paul Bogdan
UU	UU	UU	UU		19b. TELEPHONE NUMBER 213-821-5720

RPPR Final Report
as of 15-Feb-2018

Agency Code:

Proposal Number: 70347EGDRP **Agreement Number: W911NF-17-1-0076**
INVESTIGATOR(S):

Name: Paul Bogdan
Email: pbogdan@usc.edu
Phone Number: 2138215720
Principal: Y

Organization: **University of Southern California**
Address: Contracts & Grants, Los Angeles, CA 900890701
Country: USA

DUNS Number: 072933393 EIN: 951642394
Report Date: 16-Apr-2018 Date Received: 12-Jan-2018
Final Report for Period Beginning 17-Jan-2017 and Ending 16-Jan-2018
Title: Decoding the Principles of Emergence and Resiliency in Biological Collective Systems - A Multi-scale Approach
Begin Performance Period: 17-Jan-2017 **End Performance Period:** 16-Jan-2019
Report Term: 0-Other

Submitted By: Paul Bogdan Email: pbogdan@usc.edu
Phone: (213) 821-5720

Distribution Statement: 1-Approved for public release; distribution is unlimited.

STEM Degrees: 0 **STEM Participants:** 4

Major Goals: Microbial communities are known to display complex, collective behaviors. However, the underlying principles as to how these behaviors arise despite uncertainty in molecular and cellular components of the system remains unclear. Consequently, we examine the robustness of collective behaviors in microbial communities, using pattern formation of wild coliform bacteria as a model system. Many coliform bacteria naturally form complex dynamic patterns that arise from the combination of chemotaxis, nutrient degradation, and the exchange of amino acids between cells. Using both quantitative experimental methods and several theoretical frameworks, we dissect bacterial pattern formation at multiple scales, from the molecules to individual cells to self-organizing populations. By comparing pattern formation from multiple wild isolates, we attempt to identify universal principles that govern robust, collective behaviors in biological systems.
Towards this end, we adopt a multiscale approach combining experimental and theoretical approaches for the following research goals:
(1) We develop a mathematical framework for characterizing and classifying the irregularity in microbial pattern formation and validate it against experimental measurements.
(2) We determine the variability of protein copy numbers in living cells and develop a computational framework for measuring and predicting how noise in cellular components affects the overall system-level behavior.
(3) In measurements of individual cells, we analyze behaviors such as chemotactic response, signaling potential, and swimming speed to predict how single-cell heterogeneity contributes to complex, collective behavior.
(4) We develop a mathematical and experimental framework for identifying the single-cell functional states and quantify the cell-to-cell communication that lead to complex pattern formation. We define an information theoretic inspired framework for measuring how cell processing and cell-cell communication contribute to the degree of emergence, self-organization and robustness.
(5) We propose a combined mathematical and experimental framework for investigating the robustness of pattern formation when two populations of pattern forming bacteria coexist in the same space.
This project combines experimental tools including the tools of synthetic biology, fluorescence and brightfield microscopy at multiple length and time scales, and microfluidic functional assays of single-cell behavior with theoretical tools including agent-based models, non-equilibrium master equations, nonparametric statistics, systems of coupled partial differential equations, and novel analytical methods to predict and control the behavior of collective systems.
In this performance period, we did not modify or made changes in the approach or methods.

Accomplishments: In this performance period (August 1st, 2017 – January 15th, 2018), the major activities and specific objectives accomplished are as follows:

RPPR Final Report
as of 15-Feb-2018

1. We developed a mathematical framework to characterize the emergence and self-organization of microbial communities from sparse spatio-temporal time-lapse imaging data. From microbial communities to cancer cells, many such complex collectives are said to possess emergent and self-organizing behavior. However, we lack a universal and rigorous mathematical framework to quantify the degree of emergence and self-organization of biological swarms especially when considering that only sparse spatiotemporal macroscopic data is available due to technological and scientific challenges. To overcome these challenges (e.g., spatiotemporal chemoattractant concentrations, cell swimming trajectories), we propose a multi-fractal inspired framework for quantifying the degree of emergence and self-organization from time-lapse imaging with low resolution in time (several minutes between two images) and space (tens of micrometers). Emergence describes the rate of change of the probability distribution characterizing the aggregation process and can be used to detect if the emergent phenomena moves the community into a state of low energy, the case of Enterobacter cloacae, or maximum energy. We also establish mathematical connections between the proposed emergency metric and the free energy concept in statistical physics. The self-organization metric we define shows the degree of similarity (order) between different regions of the environment and the synchronization of the location of new aggregates with the previous ones. More precisely, as the experiment advances, the new Enterobacter cloacae aggregates align their location with one of the old aggregates across all the regions of the petri dish suggesting an increase in self-organization.
The benefits of this pioneering quantification strategies of emergence and self-organization consist of being able to discern when complex collectives (collective systems) display an intelligent behavior, categorize their swarming behavior with reference systems, compare and contrast them for the purpose of selecting the optimal swarms or for optimizing intelligent autonomous swarms.

2. We isolated and characterized 11 wild strains of bacteria capable of forming complex, dynamic patterns. For each strain, time-lapse revealed emergent patterns in the form of swarm rings and spots that appear due to the coordinated movement of large populations of cells. We characterized the differences in these strains through whole genome sequencing and measurements of motility and chemotaxis at the single-cell level. Genome sequencing revealed little variation in the presence and absence of the approximately 50 genes associated with microbial pattern formation, despite striking differences in the emergent patterns formed by each strain. Analysis of the single-cell swimming trajectories revealed heterogeneity in motility characteristics, including the distribution of swimming speeds, tumbling frequencies, and turning angle preferences. Surprisingly no one single-cell characteristic was correlated with variability in emergent properties such as the velocity of the swarm ring or spot distributions and sizes. These results indicate that the emergent properties of the system are not strongly determined by a single characteristic of individual agents in the system, but instead many parameters of individual agents collectively contribute to variability observed in population-level collective behavior. We have begun to develop modeling approaches to further examine the interplay between single-cell motility and chemotactic behavior and macroscale collective motion.

3. Interfacing complex collective biological systems with cyber platforms and engineering their dynamics and behavior is an open grand challenge. One such complex collective systems is represented by human brain for which developing brain machine interfaces will allow to harvest information from the (physical) brain through sensing mechanisms, extract information about the underlying processes, and decide/actuate accordingly to guide and control complex engineered systems such as swarms of aerial and ground vehicles. Nonetheless, the brain interfaces are still in their infancy, but reaching to their maturity quickly as several initiatives are released to push forward their development (e.g., NeuraLink by Elon Musk and ‘typing-bybrain’ by Facebook). State-of-the-art EEG-based non-invasive brain interfaces entail a highly skilled neuro-functional approach and evidence-based a priori knowledge about specific signal features and their interpretation from a neuro-physiological point of view. By building on mathematical concepts developed in this project, we propose new models that equip us with a fractal dynamical characterization of the brain processes. The model parameters can be seen as explainable from a system’s perspective and used for classification using machine learning algorithms and/or actuation laws obtained via control system’s theory. We evaluated our approach on real EEG-datasets to assess and validate the proposed methodology. The classification accuracies are high even with few training samples.

RPPR Final Report
as of 15-Feb-2018

Training Opportunities: This grant supported the training and professional development of the following students:

1. One graduate student, Valeriu Balaban from PI Bogdan's group working on developing mathematical approaches and codes, testing the proposed mathematical and algorithmic tools on synthetic and experimental case studies in collaboration with PI Boedicker's group, and summarizing our research results in a technical report and submitted manuscript. Valeriu Balaban has learned not only basic concepts of computer vision for processing and preparing the time lapse imaging data of microbial communities for the subsequent mathematical analysis, but also gained a solid background in fractal theory, statistical physics, information theory, nonlinear dynamics, and statistical machine learning. Working in close collaboration with PI Boedicker's group also offered unique opportunities for learning not only theoretical concepts related with synthetic and system biology, but also to test his background thanks to this project. Valeriu Balaban was in part supported by the USC fellowship.

2. Another graduate student, Xiaokan Guo from PI Boedicker's group was supported by this grant. As part of this project he developed new research skills to experimentally measure the chemotactic behavior and motility of individual bacterial cells. These experiments also enabled him to sharpen his image analysis skills, as he adapted and developed image analysis algorithms to extract single-cell parameters from movies of cell motility. He has also benefitted from interactions with the Bogdan group, introducing him to several new analytical tools and new perspectives in biophysic.

3. Another graduate student, Gaurav Gupta from PI Bogdan's group contributed to our project research discussions, developed numerous mathematical derivations in order to determine the right mathematical expressions for quantifying the degree of emergence and self-organization in microbial communities, and extended a number of mathematical ideas into new models for characterizing, modeling and analyzing the dynamics of complex collective biological systems such as the brain-in-action. Thanks to this unique project, Gaurav Gupta has learned concepts from statistical physics (free energy, entropy), fractal theory, and fractional calculus and was able to make a number of novel contributions to the field of system identification and machine learning under unknown unknowns. Gaurav Gupta was supported by this grant and by DARPA Young Faculty Award of PI Bogdan. The mathematical approaches and algorithms we developed in this project will be applied, tested and evaluated in a number of DoD problems such as the viral prediction and gene expression dynamics modeling in the DeepPurple biochronicity program.

4. This grant also has supported an undergraduate student, now a lab technician, Sean Lim. Although his salary was not directly paid from this award, funds were used for experimental supplies. He worked closely with both the Bogdan group and Xiaokan Guo. As part of this project he has learned many new laboratory skills, including single-cell characterization, microscopy, and gene sequencing. He will be attending medical school in the fall, and this experience has given him new perspectives on biophysics, experimental research, and emergent properties of biological systems.

PIs Bogdan and Boedicker interacted with the above mentioned students through one-on-one meetings as well as bi-weekly project meetings where we discussed the challenges we faced, proposed solutions and guidelines and planned our research. We have maintained a Google doc and an overleaf account to summarize our research progress and worked collaboratively on one submitted manuscript on quantifying the emergence and self-organization of microbial communities. We are currently planning our research activity to summarize our new research results and plan collaborative submissions in Spring 2018.

RPPR Final Report
as of 15-Feb-2018

Results Dissemination: During August 1st, 2017 – January 15th, 2018, the research results and accomplishments in this grant have been disseminated as follows:

1. PI Bogdan was invited and gave a Seminar Lecture entitled “Analytical Tools for Cyber-Physical Systems: Lessons Learned from Complex (Biological) Systems” in the Decision and Control Laboratory (DCL) at Georgia Tech, on November 10th, 2017.

2. The research accomplishments on mining complex dynamics of collective biological systems such as the human brain were summarizing in a manuscript that was accepted and will appear in the Proceedings of the International Conference on Cyber-Physical Systems (ICCPS) part of the CPS Week in April 2018.

3. PI Boedicker gave an invited talk at USC as part of the Women in Science and Engineering visiting day on the Biophysics of Microbial Communities.

We summarize below the citations of accepted papers, where the above mentioned research accomplishments were described (we only summarize the new publications since last submitted report):

G. Gupta, S. Pequito, and P. Bogdan, “Re-thinking EEG-based non-invasive brain interfaces - modeling and analysis” accepted to appear in the Proceedings of the International Conference on Cyber-Physical Systems (ICCPS), CPS Week, Porto, Portugal, April 2018.

Valeriu Balaban, Sean Lim, Gaurav Gupta, James Boedicker, and Paul Bogdan, “Emergence and self-organization of Enterobacter cloacae microbial communities”, submitted.

The PDF files of the publications are attached at the end of this report or will be made available as soon as we finalize the camera ready versions.

Honors and Awards: PI Bogdan has been awarded the 2017 Defense Advanced Research Projects Agency (DARPA) Young Faculty Award for research activities that spur from this project and in particular for developing mathematical models, analysis and control algorithms for time varying complex networks. The research results have been significantly enriched by the interactions with the DARPA researchers and program officers as their feedback and research questions during the DARPA review meetings and hackthons made us think and come up with new solutions that could open new fields in observability of fractal processes or quantification of emergence, self-organization and complexity not only in biology but also in social and economic / financial sciences. Unique to our research project was the close scientific supervision and interaction with our DARPA program officers that led to a new algorithm for viral prediction from the DARPA biochronicity program.

Protocol Activity Status:

Technology Transfer: We did not have any technology transfer during this reporting period.

PARTICIPANTS:

Participant Type: PD/PI
Participant: James Boedicker
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: PD/PI
Participant: Paul Bogdan Bogdan
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
International Collaboration:

International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)
Participant: Valeriu Balaban
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)
Participant: Xiaokan Guo
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Graduate Student (research assistant)
Participant: Gaurav Gupta
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Participant Type: Undergraduate Student
Participant: Sean Lim
Person Months Worked: 12.00 **Funding Support:**
Project Contribution:
International Collaboration:
International Travel:
National Academy Member: N
Other Collaborators:

Upload Report Attachment

In this section, we briefly summarize our research accomplishments during the August 2017 to January 2018 period:

A) Many complex biological systems including the human brain, biological swarms, gene regulatory networks or protein-protein interaction networks, display complex interdependent non-stationary and long-range memory dynamics. While the mathematics of such time varying complex networks currently ignores the fractal and long-range dependence characteristics, what makes their analysis even harder is that we often have available only partial observations either in space (in terms of number of state variables that we can measure or know about) and time (only few measurements can be obtained due to time, technological and economic costs). Specifically, complex networks such as the brain, whose nodes will dynamically evolve according to a fractal order dynamical model, are often observed locally. Meaning that some of the dynamics assessed by the models are not only due to the local interaction, but might be constrained by unknown sources, i.e., stimuli that are external to the network considered. Similar theoretical problems appear in system identification problems concerning the gene regulatory networks or protein-protein interaction networks in biological cells or in describing the dynamics of autonomous complex swarms subject to environmental cues.

With these challenges in mind, we developed a mathematical model that enables us to account for the existence of both the fractional dynamics and the unknown stimuli, and determined the model that best captures the local dynamics under such stimuli. Observe that this enhances the analysis of these systems once we have an additional feature (i.e., the stimuli) that can be the main driver of a possible abnormal behavior of the complex system / network. In addition, these mathematical problems are fundamental for designing efficient brain machine interfaces that aim to harvest information from the (physical) brain through sensing mechanisms, extract information about the underlying processes, and decide/actuate accordingly.

To make the discussion more concrete, we summarize in equation (1) the mathematical model:

$$\begin{aligned} \Delta^{\alpha} x[t+1] &= Ax[t] + Bu[t] \\ y[t] &= Cx[t], \end{aligned} \tag{1}$$

where $x[t]$ is the state vector characterizing the brain activity dynamics as a collection of nodes in the time-varying complex network, α is a vector of corresponding fractional-order coefficients for each node in the network and $y[t]$ is the observed signal vector. The system matrices A, B and C are spatial-coupling, input and observation matrices, respectively. The fractional differential order α can be either a real or an integer number and thus can encode either the long-range memory (long-range dependence or non-Markovian) or the short-range memory (Markovian) properties in the dynamical system model. Without loss of generality, we assume that the sensors are dedicated to each node, and therefore the observation matrix is the identity matrix. However, the proposed mathematical framework can also consider other B matrices.

The brain activity mining and interpretation process consists of: (i) Collecting the EEG data from the dedicated sensors, (ii) Estimating the model parameters from the EEG data; and (iii) Making predictions using the estimated model parameters. In this setup, the monitoring of brain activity is obtained through EEG sensing around the motor cortex of the brain. The second step is non-trivial due to the presence of unknown inputs in equation (1). With the assumption of known input matrix B as it depends on the experimental realization and fractional-order using well-known wavelet approach, we proposed an inspired Expectation-Maximization (EM) algorithm to jointly estimate the coupling matrix A and the inputs u. This EM-inspired algorithm is summarized as follows: The initial point of the algorithm $A^{(0)}$ is estimated by least squares, i.e.,

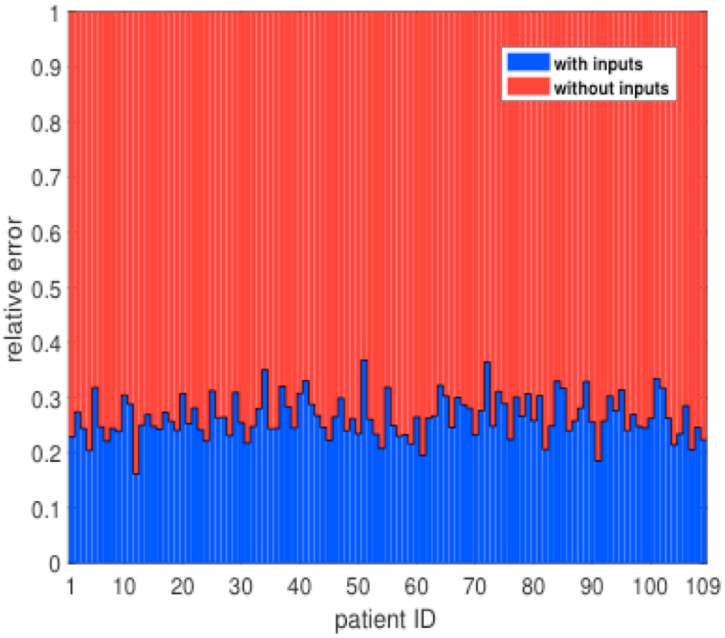


Figure 1. Time-varying complex networks subject to unknown stimuli. Error ratio when making predictions using a fractional-order dynamical model with and without inputs across 109 individuals. The fractional-order dynamical model with unknown inputs fits the EEG data much better than the case of no inputs.

$$A^{(0)} = \arg min_A \|Z - XA\|_2^2, \tag{2}$$

where Z is constructed by expanding the fractional order operator in equation (1) as $z[k] = \Delta^\alpha x[k]$ and truncating the expansion to some finite value. The expectation step of the algorithm works to estimate the unknown inputs. The unknown unknowns do not act on each sensor and are mainly sparse. Using this intuition, we have enforced the Laplace prior on the inputs u to write the E-step as

$$u[k] = \arg\min_u \left\| z[k] - A^{(l)} x[k] - Bu \right\|_2^2 + \lambda \|u\|_1, \tag{3}$$

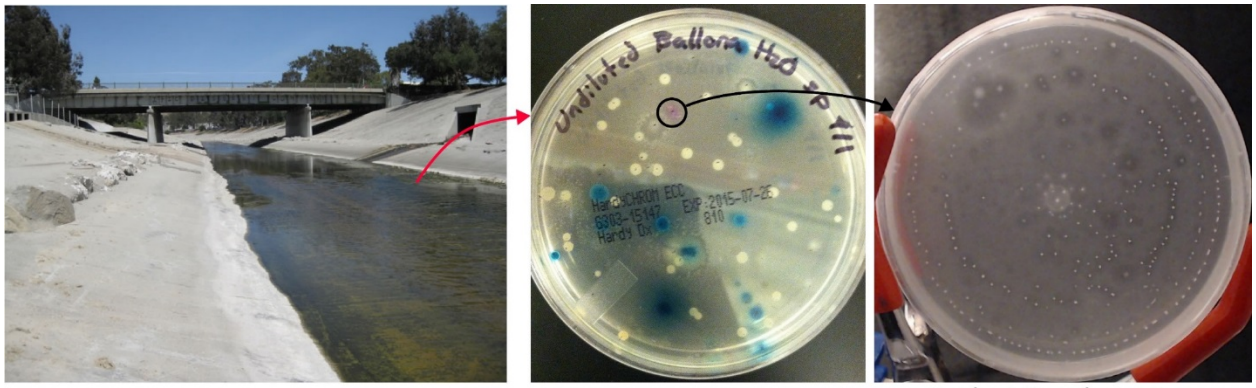
where λ is a parameter used to make a trade-off between sparsity and error. The inputs are estimated at each time point k taken into consideration. With the estimation of inputs in the E-step, the remaining part is the update of spatial coupling matrix, which is executed in the M-step as follows:

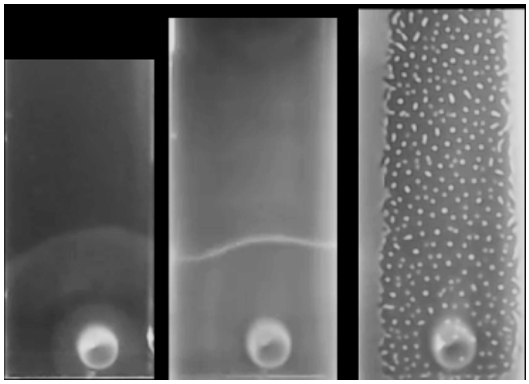
$$A^{(l+1)} = \arg\min_A \left\| \tilde{Z} - XA \right\|_2^2, \tag{4}$$

where $\hat{Z} = Z - UB$ and $U = \left[u[1], u[2], \dots, u[K]\right]^T$. We showed that the above-mentioned algorithm is convergent. It is observed that the convergence is fast for the EEG signals we analyzed and even one or two iterations are sufficient for major mean squared error reduction.

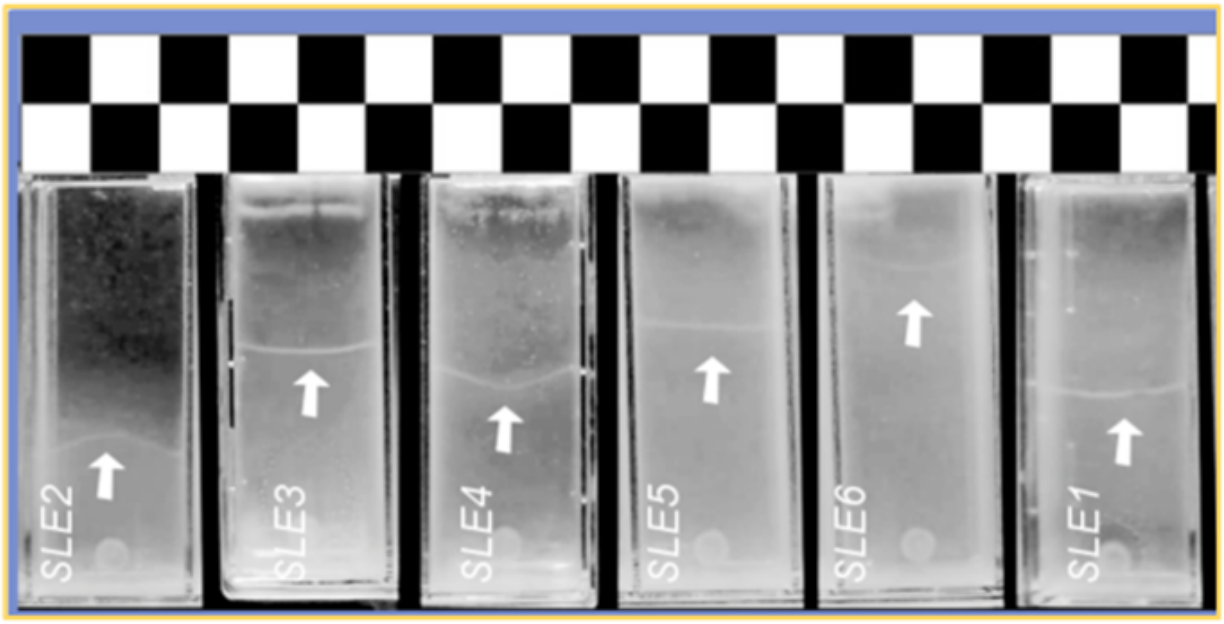
We investigated the benefits of the above-mentioned mathematical formalism by considering an EEG dataset representing the human brain activity and report the error ratio with and without inputs in Figure 1. It can be observed that the error (square root of mean squared error) for the case when the unknown inputs are included is less than one-third of the error when inputs are not considered. Moreover, the parameters of the proposed model can be seen as explainable from a system's perspective, and, subsequently, used for classification using machine learning algorithms and/or actuation laws determined using control system's theory. Besides, our proposed system identification methods and techniques have computational complexities comparable with those currently used in EEG-based brain interfaces, which enable comparable online performances. Our proposed mathematical models and algorithm are also valid using other invasive and noninvasive technologies. Of note, the classification accuracies as reported in the attached manuscript are high even on having less number of training samples.

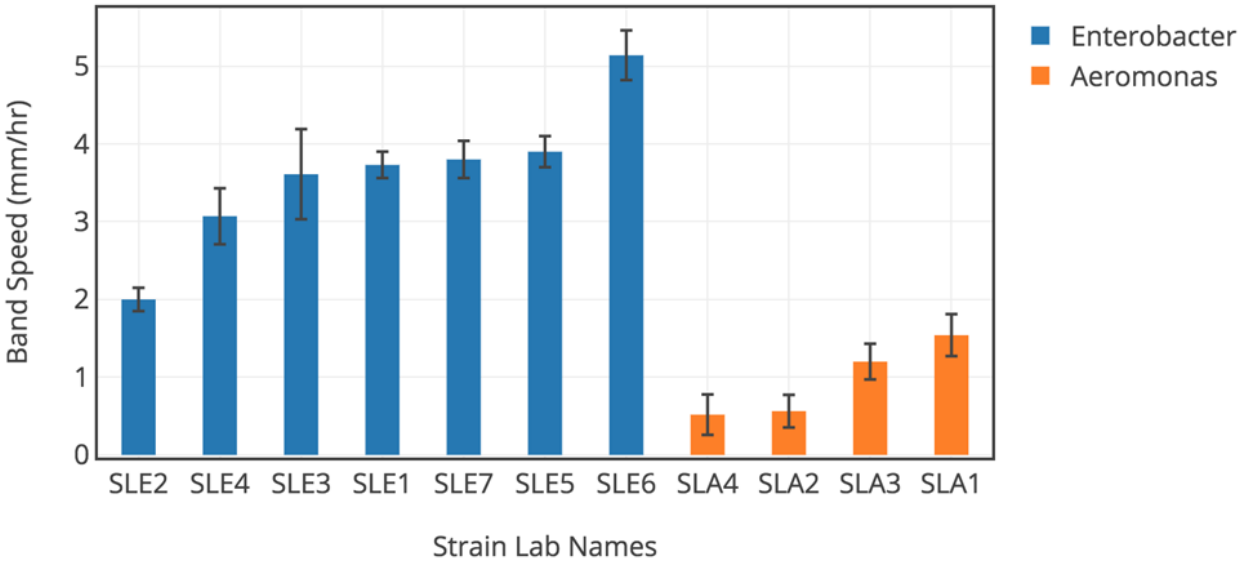
B) Wild bacterial strains were isolated from freshwater sources in Los Angeles on plates selective for bacteria related to *Escherichia coli*. Isolated strains were screened for collective pattern formation at centimeter length scales (se pictures below).



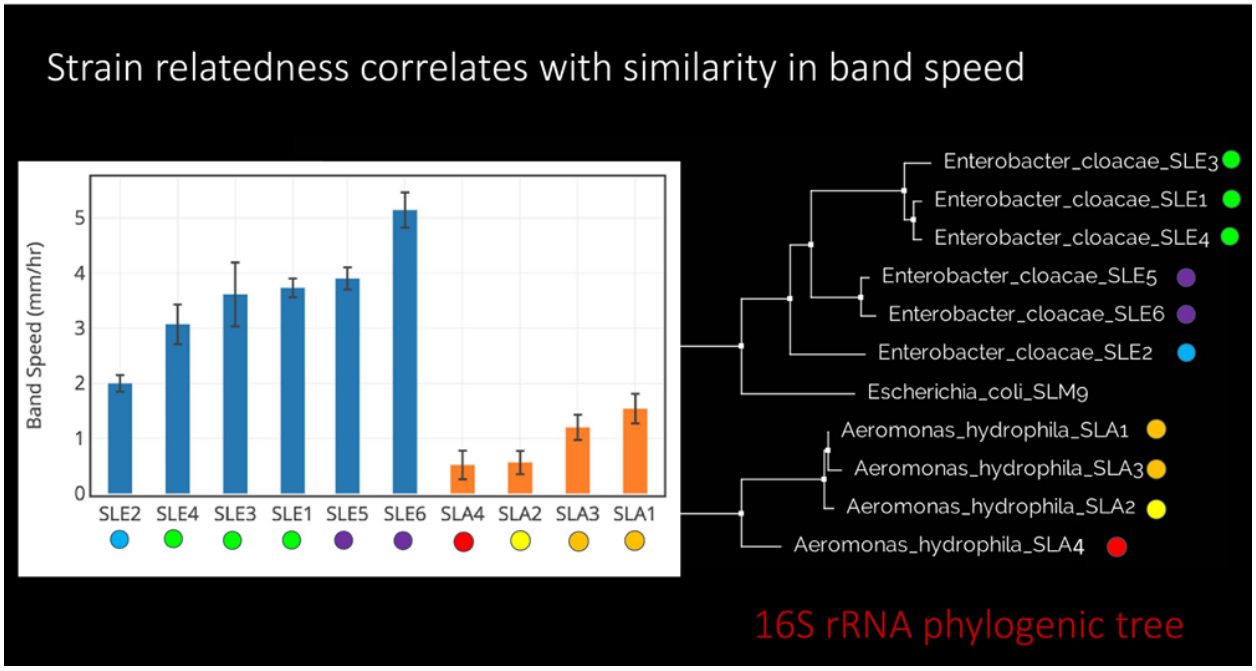


To screen for collective pattern formation, wild strains were placed at one end of a rectangular plate (see left hand side, lanes are 3 cm wide) filled with semi-soft agar. Over time, many strains formed complex patterns, including a swarm ring or a band (as shown in the middle rectangle), which radiated towards the opposite edge of the plate and spots of high cell density appeared (as shown on the right hand side panel).

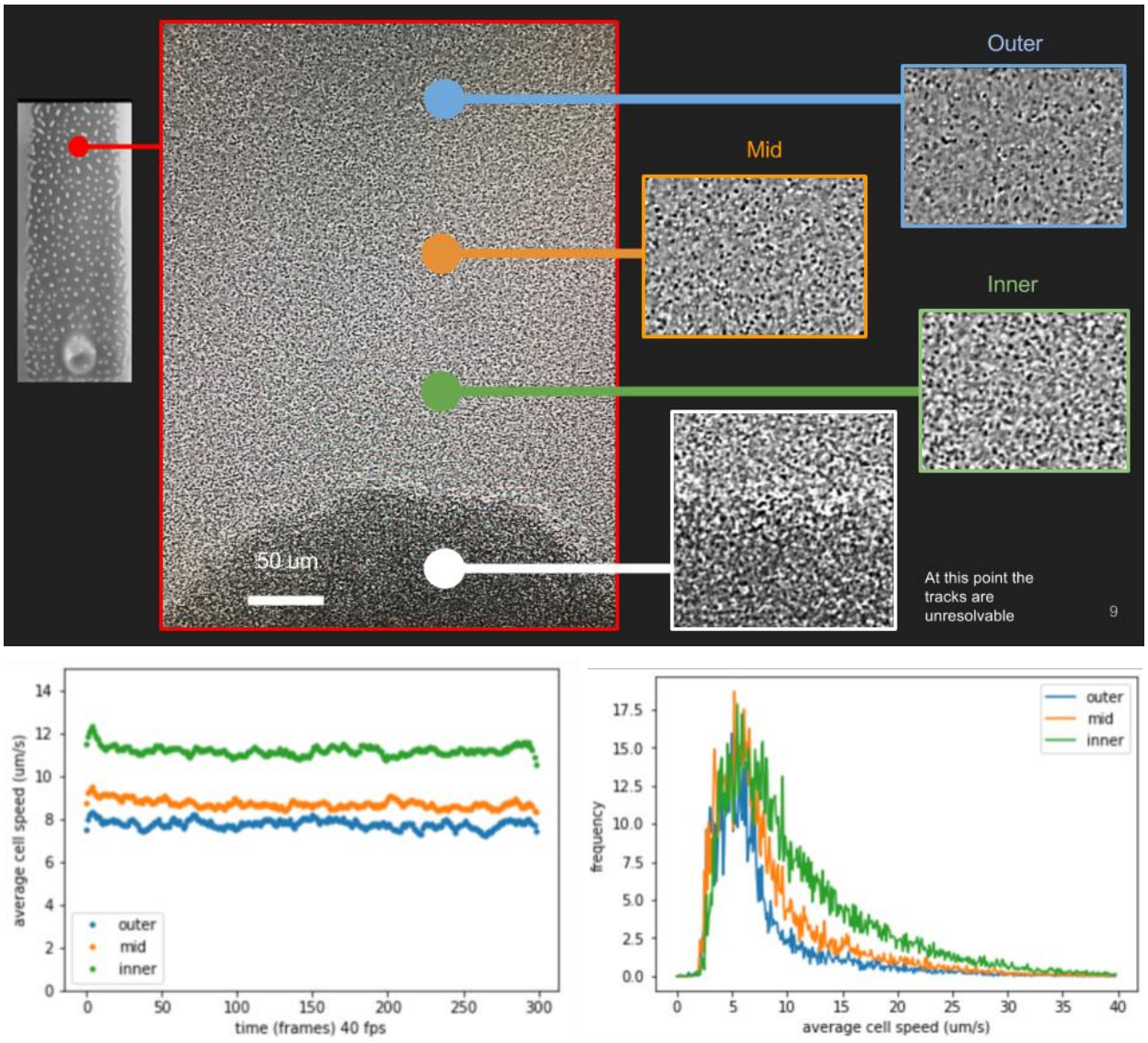




The emergent behaviors of each strain were measured, revealing variability in the collective pattern formation. We focused on two types of bacteria, seven different isolated of *Enterobacter cloacae* and four strains of *Aeromonas hydrophila*. The velocity of the high density band of cells (indicated with arrows on top image, lane width is 3 cm), was variable for the wild isolates. Subsequent analysis of these strains attempted to connect this variability in the collective dynamics of each strain with variability in the genomic content and the single-cell characteristics.

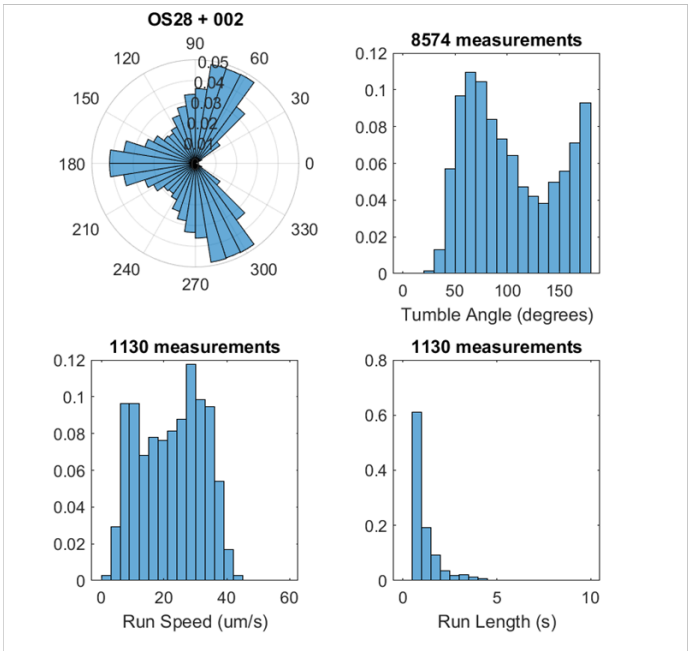


The genomes of our eleven isolates were sequenced. Alignment of the 16S rRNA sequences revealed that closely related strains had similar band speeds, suggesting that accumulated mutation in the genome modulated collective behaviors such as band formation and dynamics.

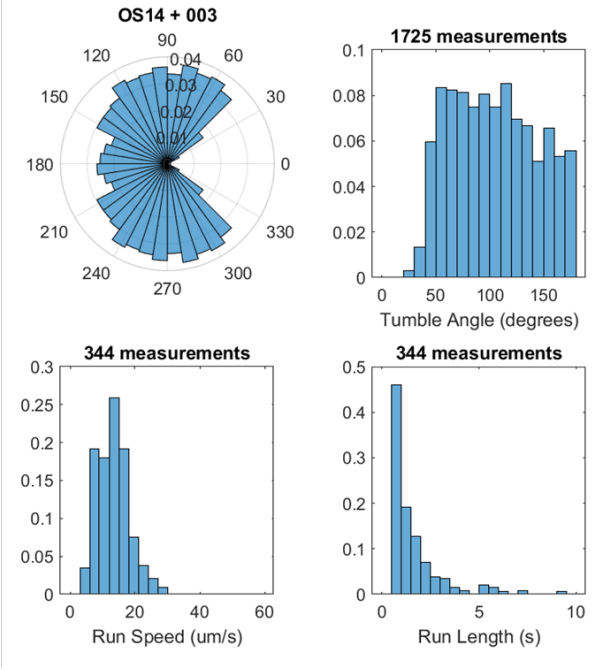


Analysis of single-cell swimming trajectories within collective patterns in the semi-soft agar assay. Cells were found to have higher velocity in the vicinity of the high cell density spots that formed on plates of *Enterobacter cloacea*.

Aeromonas hydrophila SLA2



Enterobacter cloacae SLE2



Characterization of the single-cell swimming characteristics of a wild *Enterobacter cloacae* strain and an *Aeromonas hydrophila* strain. Despite large difference in swimming speeds and tumbling angles, both strains formed similar emergent patterns.

Re-thinking EEG-based non-invasive brain interfaces: modeling and analysis

Gaurav Gupta[†] Sérgio Pequito[‡] Paul Bogdan[†]
[†]Ming Hsieh Department of Electrical Engineering, University of Southern California, CA, USA
[‡]Department of Industrial and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA
[†]{ggaurav, pbogdan}@usc.edu, [‡]goncas@rpi.edu

***Abstract*—Brain interfaces are cyber-physical systems that aim to harvest information from the (physical) brain through sensing mechanisms, extract information about the underlying processes, and decide/actuate accordingly. Nonetheless, the brain interfaces are still in their infancy, but reaching to their maturity quickly as several initiatives are released to push forward their development (e.g., NeuraLink by Elon Musk and 'typing-by-brain' by Facebook). This has motivated us to revisit the design of EEG-based non-invasive brain interfaces. Specifically, current methodologies entail a highly skilled neuro-functional approach and evidence-based *a priori* knowledge about specific signal features and their interpretation from a neuro-physiological point of view. Hereafter, we propose to demystify such approaches, as we propose to leverage new models that equip us with a fractal dynamical characterization of the underlying processes. Subsequently, the parameters of the proposed models can be seen as explainable from a system's perspective, and, subsequently, used for classification using machine learning algorithms and/or actuation laws determined using control system's theory. Besides, the proposed system identification methods and techniques have computational complexities comparable with those currently used in EEG-based brain interfaces, which enable comparable online performances. Furthermore, we foresee that the proposed models and approaches are also valid using other invasive and non-invasive technologies. Finally, we illustrate and experimentally evaluate this approach on real EEG-datasets to assess and validate the proposed methodology. The classification accuracies are high even on having less number of training samples.**

***Index Terms*—brain interfaces, spatiotemporal, fractional dynamics, unknown inputs, classification, motor prediction**

I. Introduction

We have recently testimony a renewed interest in invasive and non-invasive brain interfaces. Elon Musk has released the NeuraLink initiative [1] that aims to develop devices and mechanisms to interact with the brain in a symbiotic fashion, thus merging the artificial intelligence with the human brain. The potential is enormous since it would ideally present a leap in our understanding of the brain, and an unseen enhancement of its functionality. Alternatively, Facebook just announced the 'Typing-by-Brain' project [2] that gathered a team of 60 researchers whose target is to be capable of writing 100 words per minute that contrasts with the state-of-the-art of 0.3 to 0.82 words per minute assuming an average of 5 letters per word. Towards this goal, Facebook has invested in developing new non-invasive optical imaging technology that is five times faster and portable with respect to the one available on the market and would possess increased spatial and temporal resolution. Nonetheless, these are just some of the initiatives among others by some big Silicon Valley players that want to commercialize brain technologies [3].

Part of the motivation for the 'hype' in the use of neurotechnologies – both invasive and non-invasive brain interfaces – is due to their promising application domains [4]: (*i*) *replace*, i.e., the interaction of the brain with a wheelchair or a prosthetic device to replace a permanent functionality loss, (*ii*) *restore*, i.e., to bring to its normal use some reversible loss of functionality such as walking after a severe car accident or limb movement after a stroke; (*iii*) *enhance*, i.e., to enable to outperform in a specific function or task, as for instance an alert system to drive for long periods of time while attention up; and (*iv*) *supplement* as in equipping one with extra functionality, as a third arm to be used during surgery. Notwithstanding, these are just some of the (known) potential uses of neurotechnology.

Despite the developments and promise of future applications of brain interfaces (some of which we cannot currently conceive), we believe that current approaches to both invasive and non-invasive brain interfaces can greatly benefit from cyber-physical systems (CPS) oriented approaches and tools to increase their efficacy and resilience. Hereafter, we propose to focus on non-invasive technology relying on electroencephalogram (EEG) and revisit it through a CPS lens. Yet, we believe that the proposed methodology can be easily applicable to other technologies, e.g., electromagnetic fields (magnetoencephalography (MEG) [5], and the hemodynamic responses associated to neural activity (e.g. functional magnetic resonance imaging (fMRI) [6], [7], and functional near-infrared spectroscopy (fNIRS) [8]). Nonetheless, these technologies present several drawbacks compared to EEG, e.g., cost, scalability, and user comfort, which lead us to focus on EEG-based technologies. Similar argument can be applied in the context of non-invasive versus invasive technologies that require patient surgery.

Traditional approach to EEG neuroadaptive technology consists of proceeding through the following steps [9], [10]: (*a*) signal acquisition (in our case, measurement of the EEG time-series); (*b*) signal processing (e.g., filtering with respect to known error sources); (*c*) feature extraction (i.e., an artificial construction of the signal that aims to capture quantities of interest); (*d*) feature translation (i.e., classification of the signal according to some plausible neurophysiological hypothesis);

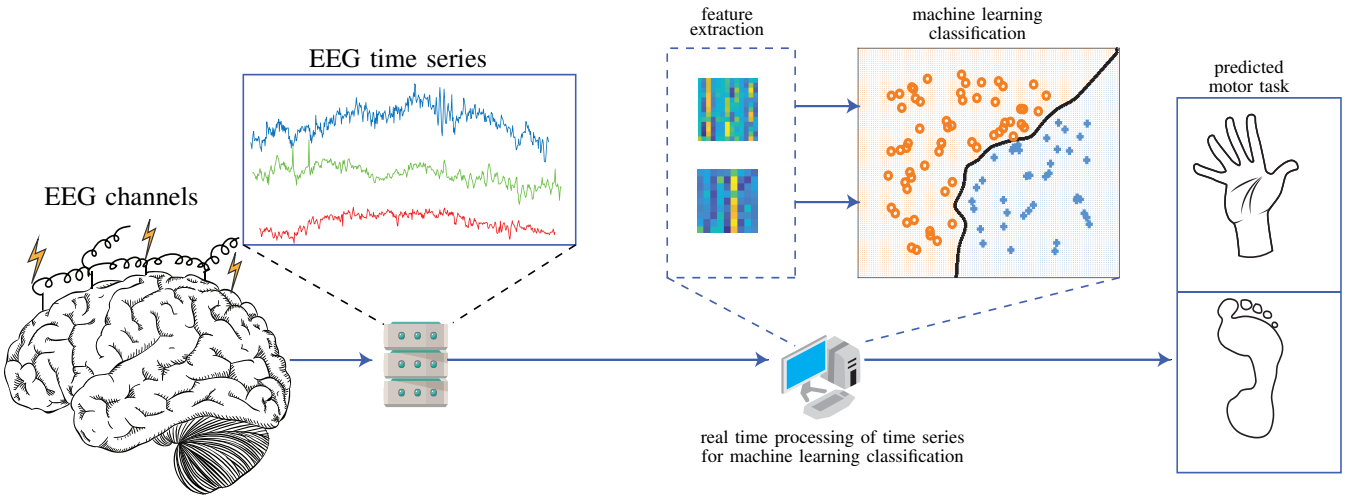


Fig. 1: A systematic process flow of the Brain interface. The imagined motor movements of the subject are captured in the form of EEG time series which are then fed to the computational unit. A fractional-order dynamic model is estimated for the time series and the model parameters are used as features for machine learning classification. The output of classifier predicts various motor movements with certain confidence.

and (*e*) decision making (e.g., provide instructions to the computer to move a cursor, or a wheelchair to move forward) – see also Figure 1 for an overview diagram.

In this paper, we propose to merge steps (b) and (c) motivated by the fact that these often accounts for spatial or temporal, and are only artificially combined in a later phase of the pipeline, i.e., at step (d) of feature translation. Thus, we argue that the previous approach discards several spatial-temporal properties that can be weighted for signal processing and feature extraction phases. In other words, current EEG brain interfaces require one to have an understanding of the different regions of the brain responsible, for instance, for motor or visual actions, as well as artificial frequency bands that are believed to be more significant for a specific action (also known as *evidence-based*). Besides, one needs to understand and anticipate the most likely causes noise/artifacts in the EEG data collected and filter out entire frequency bands, which possibly compromises phenomena of interest not being available for post-processing. Instead, we propose a modeling capability to enable the modeling of long-range memory time-series that at the same time accounts for unknown stimuli, e.g., artifacts or inter-region communication.

A. Related Work and Novel Contributions

By being able to properly model EEG time-series with models that account for realistic setups, brain interfaces methods, which are mainly detectors can be transformed into decoders. In other words, we do not want to solely look for the existence of a peak of activity in a given band that is believed to be associated with a specific action, but we want to decompose the signal into different features, i.e., parameters of our model, that are interpretable. Thus, allowing us to understand how different regions communicate during a specific action/task, as well as the external stimuli driving the process and the different time-scales at which the underlying process occurs. In engineering, this will enable us to depart from a skill dependent situation to general context analysis, which will enhance the resilience of the approaches for practical non-surgical brain interfaces. Besides, it will equip bioengineers, neuroscientists, and physicians with an exploratory tool to pursue new technologies for neuro-related diagnostics and treatments, as well as neuro-enhancement.

The proposed approach departs from those available in the literature, see [4], [9], [10] and references therein. In fact, to the best of authors' knowledge, in the context of noninvasive EEG-based technologies, [11] is the only existing work that explores fractional-order models in the context of single-channel analysis, which contrasts with the spatiotemporal modeled leveraged in this paper that copes with multi-channel analysis. Furthermore, the methodology presented in this paper also accommodate unknown stimuli [12]. For which efficient algorithms are proposed and leveraged hereafter to simultaneously retrieve the best model that conforms with unknown stimuli, and separating the unknown stimuli from the time-series associated with brain activity. Our methods are as computationally efficient and stable as least-squares and spectral analysis methods used in a spatial and temporal analysis, respectively; thus, suitable for online implementation in nonsurgical brain interfaces.

The main contributions of the present paper are those of leveraging some of the recently proposed methods to develop new modeling capabilities for the EEG based neuro-wearables that are capable of enhancing the signal quality and decision-making. Furthermore, the parametric description of these models provides us with new features that are biologically motivated and easier to translate in the context

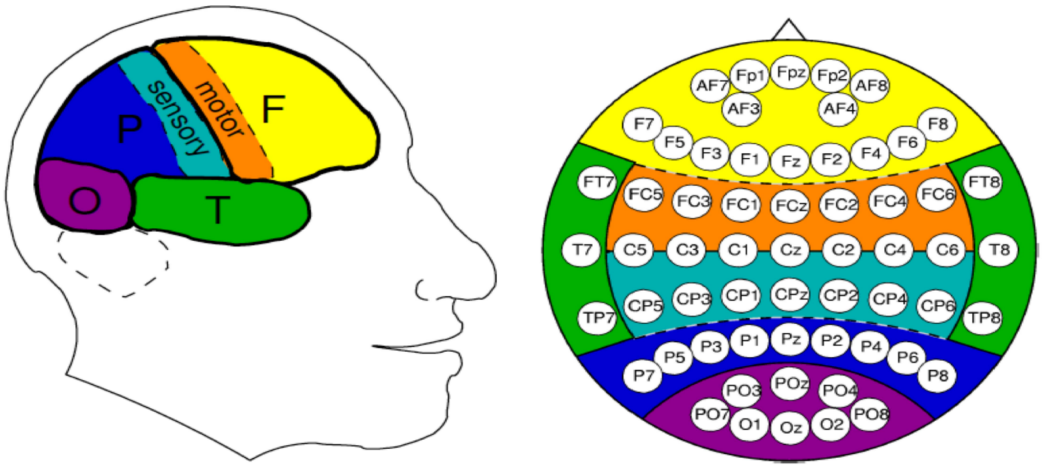


Fig. 2: Description of the brain functional regions and their corresponding location with respect to the EEG sensor cap. [image credits] Brain-Computer Interfacing at the Bernstein Focus: Neurotechnology

of brain function associated with a signal characterization, and free of signal artifacts. Thus, making the brain-related activity interpretable, which leads to resilient and functional nonsurgical brain interfaces.

B. Paper Organization

The remaining of the paper is organized as follows. Section II introduces the model considered in this paper and the main problem studied in this manuscript. Also we will see the description of the employed method for feature selection and then classification techniques. In Section III, we present an elaborated study on the datasets taken from the BCI competition [13].

II. Re-thinking EEG-based non-invasive brain interfaces

Brain interfaces aim to address the following problem.

Is it possible to classify a specific cognitive state, e.g., motor task or its imagination, by using measurements collected with a specific sensing technology that harvest information about brain activity?

In the current manuscript, we want to revisit this problem in the context of brain-computer interfaces (BCI), when dealing with EEG-based noninvasive brain interfaces. Towards this goal, we review the adopted procedure for solving this problem (see Fig. 1 for an overview), and proposed a systems' perspective that enables to enhance the BCI reliability and resilience. Therefore, in Section II-A we provide a brief overview of the EEG-based technology and the connection with the brain-areas' function associated with studies conducted in the past. Next, in Section II-B, we introduce the spatiotemporal fractional model under unknown stimuli, which will be the core of the proposed approach in this manuscript to retrieve new features for classification, and, subsequently, enhancing brain interfaces capabilities. In Section II-C, we describe how to determine the system model's parameters, and in Section II-D we describe how these can then be interpreted for classification.

A. EEG-based Technology for Brain Interfaces: a brief overview

EEG enables the electrophysiological monitoring of space-averaged synaptic source activity from millions of neurons occurring at the neocortex level. EEG has a poor spatial resolution but high temporal resolution, since the electrical activity generated at the ensemble of neurons level arrives at the recording sites within milliseconds. The electrodes (i.e., sensor) are placed over an area of the brain of interest, being the most common the visual, motor, sensory, and prefrontal cortices. Usually, they follow standard *montages* – the International 10-20 system is depicted in Fig. 2.

Most of the activity captured by the EEG electrodes is due to the interactions between inhibitory interneurons and excitatory pyramidal cells, which produces rhythmic fluctuations commonly referred to as *oscilations*. The mechanisms that generate those oscillations is not yet completely understood, but it has been already identified that some 'natural oscillations' provide evidence of activity being 'processed' in certain regions of the brain at certain 'frequencies.' Therefore, oscillatory behavior of human brain is often partitioned in bands (covering a wide range of frequencies decaying as $1/f$ in power): (*i*) δ-band (0.5-3Hz); (*ii*) θ-band (3.5-7Hz); (*iii*) α-band (8-13Hz); (*iv*) β-band (14-30Hz); and (*v*) γ-band (30-70Hz). Furthermore, several there has been some evidence that activity in certain bands is associated with sensory registration, perception, movement and cognitive processes related to attention, learning and memory [14]–[16]. Notwithstanding, such associations are often made using correlation and/or coherence techniques that only capture relationships between specific channels but are not able to assess the causality between signals that enables forecasting on the signal evolution, captured by a model-based representation as we propose to do hereafter.

Different changes in the signals across different bands are also used to interpret the event-related potentials (ERPs) in the EEG signals, i.e., variations due to specific events – see [4] for

detailed analysis. In the context of sensory-motor data used in the current manuscript to validate the proposed methodology, sensorimotor rhythms (SMRs) are often considered. These represent oscillations that are recorded over the posterior frontal and anterior parietal areas of the brain, i.e., over the sensorimotor cortices (see Fig. 2). SMRs occur mainly in the α-band (for sensors located on the top of the motor cortices), and on beta and lower gamma for those on the sensorimotor cortices [17]. Consequently, these have been used as a *default feature* for classification of motor-related execution and only the imagination of performing a motor task. Notwithstanding, the spatiotemporal modeling is not simultaneously captured through direct state-space modeling that enables the system's understanding of the dynamics of the underlying process, and, subsequently, a new set of features that can be used to improve feature translation, i.e., classification.

B. Spatiotemporal Fractional Model under Unknown Stimuli

A multitude of complex systems exhibits long-range (non-local) properties, interactions and/or dependencies (e.g., power-law decays in memories). Example of such systems includes Hamiltonian systems, where memory is the result of stickiness of trajectories in time to the islands of regular motion [18]. Alternatively, it has been rigorously confirmed that viscoelastic properties are typical for a wide variety of biological entities like stem cells, liver, pancreas, heart valve, brain, muscles [18]–[26], suggesting that memories of these systems obey the power law distributions. These dynamical systems can be characterized by the well-established mathematical theory of fractional calculus [27], and the corresponding systems could be described by fractional differential equations [28]–[32]. However, it is until recently that fractional order system (FOS) starts to find its strong position in a wide spectrum of applications in different domains due to the availability of computing and data acquisition methods to evaluate its efficacy in terms of capturing the underlying system states evolution.

Subsequently, we consider a linear discrete time fractional-order dynamical model under unknown stimuli (i.e., inputs) described as follows:

$$\begin{aligned} \Delta^{\alpha} x[k+1] &= Ax[k] + Bu[k] \\ y[k] &= Cx[k], \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$ is the state, $u \in \mathbb{R}^p$ is the unknown input and $y \in \mathbb{R}^n$ is the output vector. Also, we can describe the system by its matrices tuple (α, A, B, C) of appropriate dimensions. The coupling matrix A represents the spatial coupling between the states across time while the input coupling matrix B determines how inputs are affecting the states. In what follows, we assume that the input size is always strictly less than the size of state vector, i.e., $p < n$. The difference between a classic linear time-invariant and the above model is the inclusion of fractional-order derivative whose expansion and discretization [33] for any ith state $(1 \leqslant i \leqslant n)$ can be written as

$$\Delta^{\alpha_i} x_i[k] = \sum_{j=0}^{k} \psi(\alpha_i, j) x_i[k-j], \tag{2}$$

where α_i is the fractional order corresponding to the ith state and $\psi(\alpha_i, j) = \frac{\Gamma(j-\alpha_i)}{\Gamma(-\alpha_i)\Gamma(j+1)}$ with $\Gamma(.)$ denoting the gamma function.

Having defined the system model, the system identification, i.e., estimation of model parameters, from the given data is an important step. It becomes nontrivial when we have unknown inputs since one has to be able to differentiate which part of the evolution of the system is due to its intrinsic dynamics and what is due to the unknown inputs. Subsequently, the analysis part that we need to address is that of system identification from the data, as described next.

C. Data driven system identification

The problem consists of estimating α, A and inputs $\{u[k]\}_t^{t+T-2}$ from the given limited observations $y[k]$, $k = [t, t+T-1]$, which due to the dedicated nature of sensing mechanism is same as $x[k]$ and under the assumption that the input matrix B is known. The realization of B can be application dependent and is computed separately using experimental data – as we explore later in the case study, see Section III. For the simplicity of notation, let us denote $z[k] = \Delta^{\alpha} x[k+1]$ with k chosen appropriately. The pre-factors in the summation in (2) grows as $\psi(\alpha_i, j) \sim \mathcal{O}(j^{-\alpha_i - 1})$ and, therefore, for the purpose of computational ease we would be limiting the summation in (2) to J values, where $J > 0$ is sufficiently large. Therefore, $z_i[k]$ can be written as

$$z_i[k] = \sum_{j=0}^{J-1} \psi(\alpha_i, j) x[k+1-j], \tag{3}$$

with the assumption that $x[k], u[k] = 0$ for $k \leqslant t-1$. Using the above introduced notations and the model definition in (1), the given observations can be written as

$$z[k] = Ax[k] + Bu[k] + e[k], \tag{4}$$

where $e \sim \mathcal{N}(0, \Sigma)$ is assumed to be Gaussian noise independent across space and time. For simplicity we would assume that $\Sigma = \sigma^2 I$. Also, let us denote the system matrices as $A = [a_1, a_2, \ldots, a_n]^T$ and $B = [b_1, b_2, \ldots, b_n]^T$. The vertical concatenated states and inputs during an arbitrary window of time as $X_{[t-1,t+T-2]} = [x[t-1], x[t], \ldots, x[t+T-2]]^T$, $U_{[t-1,t+T-2]} = [u[t-1], u[t], \ldots, u[t+T-2]]^T$ respectively, and for any ith state we have $Z_{i,[t-1,t+T-2]} = [z_i[t-1], z_i[t], \ldots, z_i[t+T-2]]^T$. For the sake of brevity we would be dropping the time horizon subscript from the above matrices as it is clear from the context.

Since the problem of joint estimation of the different parameters is highly nonlinear, we proceed as follows: (*i*) we estimate the fractional order α using the wavelet technique described in [34]; and (*ii*) with α known, the z in equation (3) can be computed under the additional assumption that the system matrix B is known. Therefore, the problem now

reduces to estimate A and the inputs $\{u[k]\}_t^{t+T-2}$. Towards this goal, we borrow the expectation-maximization (EM) [35] like algorithm from [12]. Briefly, the EM algorithm is used for maximum likelihood estimation (MLE) of parameters subject to hidden variables. Intuitively, in our case, in Algorithm 1, we estimate A in the presence of hidden variables or *unknown unknowns* $\{u[k]\}_t^{t+T-2}$. Therefore, the 'E-step' is performed to average out the effects of unknown unknowns and obtain an estimate of u, where due to the diversity of solutions, we control the sparsity of the inputs (using the parameter λ'). Subsequently, the 'M-step' can then accomplish MLE estimation to obtain an estimate of A.

It was shown theoretically in [12] that the algorithm is convergent in the likelihood sense. It should also be noted that the EM algorithm can converge to saddle points as exemplified in [35]. The Algorithm 1 being iterative is crucially dependent on the initial condition for the convergence. We will see in the Section III that the convergence is very fast making it suitable for online estimation of parameters.

Algorithm 1: EM algorithm

Input: $x[k], k \in [t, t+T-1]$ and B
Output: A and $\{u[k]\}_t^{t+T-2}$
Initialize compute α using [34] and then $z[k]$. For $l = 0$, initialize $A^{(l)}$ as

$$a_i^{(l)} = \arg\min_a ||Z_i - Xa||_2^2$$

repeat
(i) **'E-step'**: For $k \in [t, t+T-2]$ obtain $u[k]$ as

$$u[k] = \arg\min_u ||z[k] - A^{(l)}x[k] - Bu||_2^2 + \lambda'||u||_1,$$

where $\lambda' = 2\sigma^2\lambda$;
(ii) **'M-step'**:
obtain $A^{(l+1)} = [a_1^{(l+1)}, a_2^{(l+1)}, \ldots, a_n^{(l+1)}]^T$ where

$$a_i^{(l+1)} = \arg\min_a ||\tilde{Z}_i - Xa||_2^2,$$

and $\tilde{Z}_i = Z_i - Ub_i$;
$l \leftarrow l + 1$;
until until converge;

D. Feature Translation (Classification)

The EEG signals directly from the sensors although carrying vital information may not be directly useful for making the predictions. However, by representing the signals in terms of parametric model (α, A) and the unknown signals as we did in the last section, we can gain better insights. The parameters of the model being representative of the original signal itself can be used to make a concise differentiation.

The A matrix represents how strong is the particular signal and how much it is affecting/being affected by the other signals that are considered together. While performing or imagining particular motor tasks, certain regions of the brain gets more activated than others. Therefore, the columns of A which represent the coefficients of the strength of a signal affecting other signals can be used as a feature for classification of motor tasks. In this work, we will be considering the machine learning based classification techniques like logistic regression and Support Vector Machines (SVM) [36]. The choice of kernels would vary from simple 'linear' to radial basis function (RBF), i.e., $k(x_i, x_j) = e^{-\gamma(x_i - x_j)^2}$. The value of parameters of the classifier and possibly of the kernels would be determined using cross-validation. The range of parameters in the cross-validation would be from $2^{-5}, \ldots, 2^{15}$ for γ and $2^{-15}, \ldots, 2^3$ for $C = 1/\lambda$, both in the logarithmic scale, where λ is the regularization parameter which appears in optimization cost of the classifiers [36].

III. Case Study

We will now illustrate the usefulness of the fractional-order dynamic model with unknown inputs in the context of classification for Brain Computer Interface (BCI). We have considered two datasets from the BCI competition [13]. The datasets were selected on the priority of larger number of EEG channels and number of trials for training. The available data is split into the ratio of 60% and 40% for the purpose of training and testing, respectively.

A. Dataset-I

We consider for validation the dataset labeled 'dataset IVa' from BCI Competition-III [37]. The recording was made using BrainAmp amplifiers and a 128 channel electrode cap and out of which 118 channels were used. The signals were band-pass filtered between 0.05 and 200 Hz and then digitized at 1000 Hz. For the purpose of this study we have used the downsampled version at 100 Hz. The dataset for subject ID 'al' is considered, and it contains 280 trials. The subject was

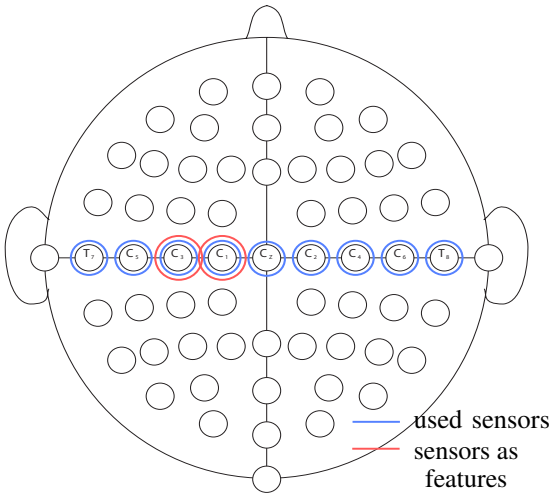
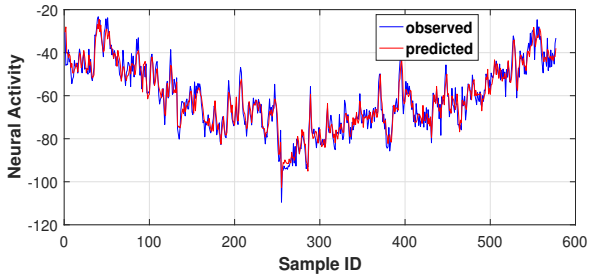


Fig. 3: A description of the sensor distribution for the measurement of EEG. The channel labels for the selected sensors are shown for dataset-I.

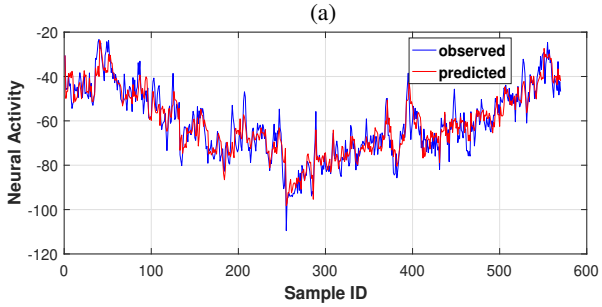
provided a visual cue, and immediately after asked to imagine two motor tasks: (R) right hand, and (F) right foot.

1) Sensor Selection and Modeling: To avoid the curse-of-dimensionality, instead of considering 118 sensors available, which implies the use of 118×118 dynamics entries for classification, only a subset of 9 sensors was considered. Specifically, only the sensors indicated in Fig. 3 are selected on the basis that only hand and feet movements need to be predicted, and only a 9×9 dynamics matrix and 9 fractional order coefficents are required for modeling the fractional order system. Besides, these sensors were selected because they are close to the region of the brain known to be associated with motor actions.

2) System Identification and Validation: The model parameters (α, A) and the unknown inputs are estimated by using the Algorithm 1. As mentioned before, the performance of the algorithm being iterative is dependent on the choice of the initial conditions. For the current case, we have observed that the algorithm converges very fast, and even a single iteration is enough. This shows that the choice of initial conditions are fairly good. The one step and five step prediction of the estimated model is shown in Fig. 4. It can be seen that the predicted values for one step very closely follow the actual values. There are some differences between the actual and predicted values for five step prediction.



(a)



(b)

Fig. 4: Comparison of predicted EEG state for the channel T_7 using fractional-order dynamical model with unknown inputs. The one step and five step predictions are shown in (a) and (b) respectively.

3) Discussion of the results: The most popular features used in the motor-imagery based BCI classification relies on exploiting the spectral information. The features used are the band-power which quantifies the energy of the EEG signals

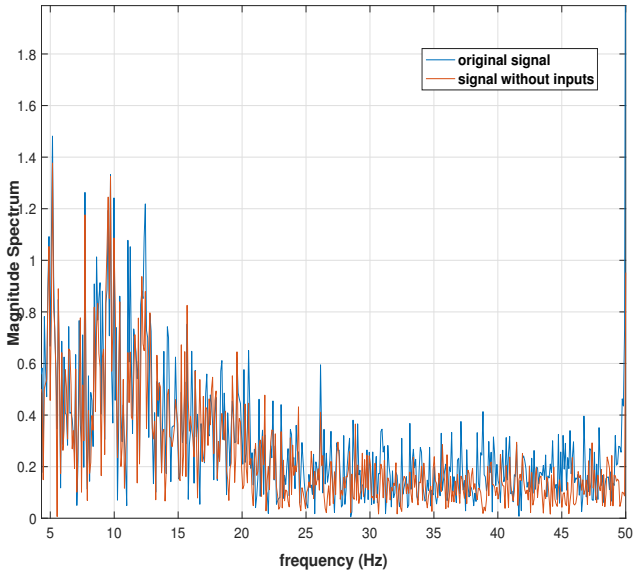


Fig. 5: Magnitude spectrum of the signal recorded by channel T_7 with and without unknown inputs.

in certain spectrum bands [38]–[40]. The motor cortex of the brain is known to be affecting the energy in the bands namely, α and β as discussed in the Section II. While it happens that an unwanted signal energy is captured in these bands as well while performing the experiments, for example neck movement, other muscle activities etc. The filtering of these so called 'unwanted' components from the original signal is a challenging task using the spectral techniques as they often share the same band.

We have taken a different approach to deal with these unknown unknowns in Section II. The magnitude spectrum of the original EEG signal and on removing the estimated unknown inputs is shown in Fig. 5. It should be observed that the original signal and the signal upon removing the unknown inputs have significant energy in the α and β bands. The unknown inputs behave somewhat like white noise which is evident from their Gaussian probability distribution (PDF) as shown in Fig. 6. The inputs are not mean zero but their PDF is centered around a mean value of approximately 58.

The model parameters (α, A) are jointly estimated with the unknown inputs using Algorithm 1, therefore the effect of the inputs is inherently taken care of in the parameters. The structure of matrix A for two different labels is shown in Fig. 7. We will be using the sensors C_3 and C_1 which are indexed as 3 and 4, respectively in the Fig. 7. We can observe that the columns corresponding to these sensors are having varied activity and hence deem to be fair candidates for the features to be used in classification. Therefore, the total number of features are going to be $2 \times 9 = 18$.

B. Dataset-II

A 118 channel EEG data from BCI Competition-III, labeled as 'dataset IVb' is taken [37]. The data acquisition technique and sampling frequencies are same as in dataset of the previous

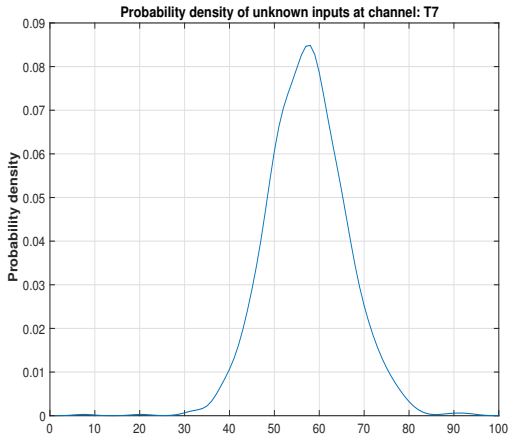


Fig. 6: Probability density function of the unknown inputs estimated from the signal recorded by channel T_7.

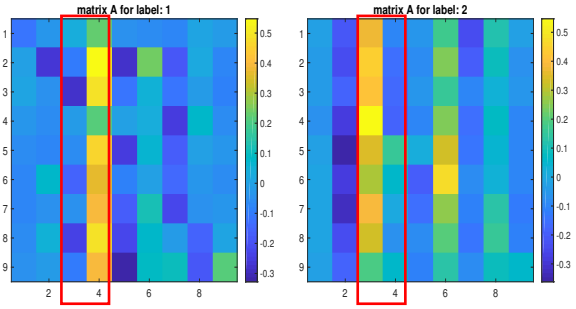


Fig. 7: Estimated A matrix of size 9×9 for the dataset-I with marked columns corresponding to the sensor index 3 and 4 used for classification.

subsection. The total number of labeled trials are 210. The subjects upon provided visual cues were asked to imagine two motor tasks, namely (L) left hand and (F) right foot.

1) Sensor Selection and Modeling: Due to the small number of training examples, we would again resort to be selecting a subset of sensors for the model estimation as we did for the dataset-I in the previous section. Since the motor tasks were left hand and feet, therefore we have selected the sensors in the right half of the brain and close to the region which is known to be associated with hand and feet movements as shown in Fig. 8. We will see in the final part of this section that selecting sensors based on such analogy helps not only in reducing the number of features, but also to gain better and meaningful results.

2) System Identification and Validation: After performing the estimation of the model (α, A) and the unknown inputs using the subset of sensors, we can see the similar performance of the model on dataset-II as was in dataset-I. The one step and five step predictions are shown in Fig. 9. The model prediction follows closely the original signal.

3) Discussion of the results: The spectrum of the original EEG signal at channel CFC_2 and its version with unknown inputs removed are shown in Fig. 10. The spectrum shows

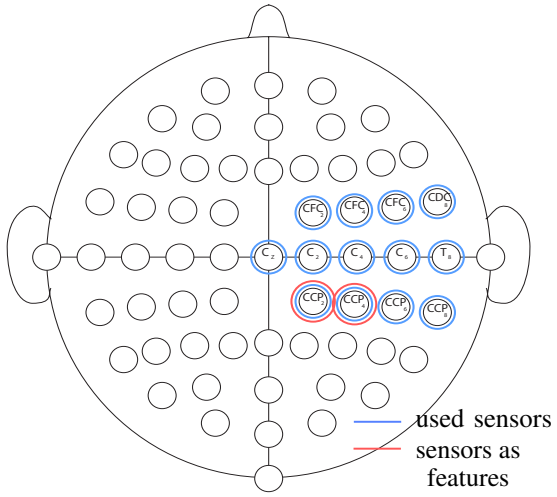
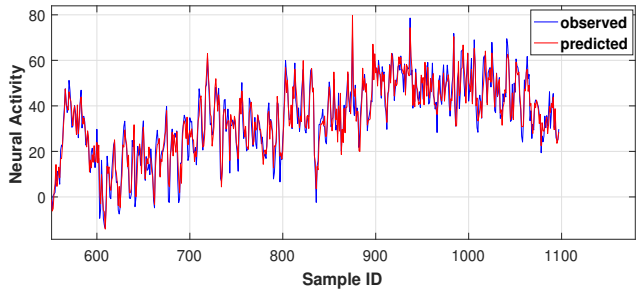
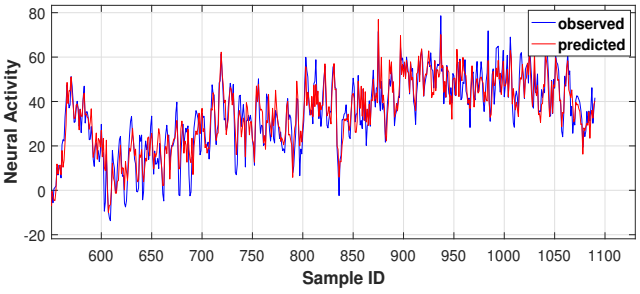


Fig. 8: A description of the sensor distribution for the measurement of EEG. The channel labels for the selected sensors are shown for dataset-II.



(a)



(b)

Fig. 9: Comparison of predicted EEG state for the channel CFC_2 using fractional-order dynamical model with unknown inputs. The one step and five step predictions are shown in (a) and (b) respectively.

peaks in the α and β bands. We can again make the similar observation as before that both of the signals share the same band and hence making it difficult to remove the effects of the unwanted inputs. The unknown inputs can be viewed as white noise and the PDF can be seen as Gaussian distributed with mean centered at around 48.

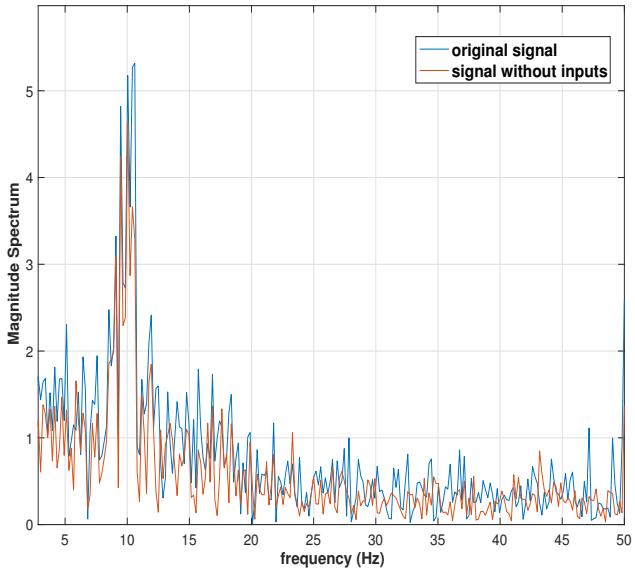


Fig. 10: Magnitude spectrum of the signal recorded by channel CFC_2 with and without unknown inputs.

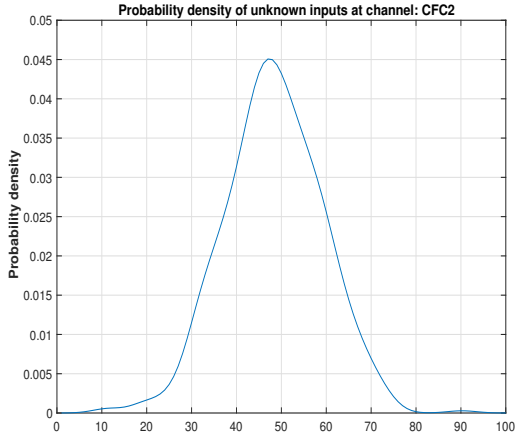


Fig. 11: Probability density function of the unknown inputs estimated from the signal recorded by channel CFC_2.

The estimated A matrix from Algorithm 1 is shown in Fig. 12 for two different labels. Out of all 13 sensors, the sensors CCP_2 and CCP_4 which are indexed as 10 and 11 in the matrix have striking different activity. The columns corresponding to these two sensors seem good choice for being the features for classification. Therefore, the total number of features are going to be $2 \times 13 = 26$ for this dataset. Next, we are going to discuss the classification accuracy for both the datasets.

C. Classification Performance

Finally, the performance of the classifier using the features explained for both the datasets can be seen in Fig. 13. The classifiers are arranged in the order of complexity from left to right with logistic regression (lR) and linear kernel being simplest and SVM with RBF kernel being most complex. We

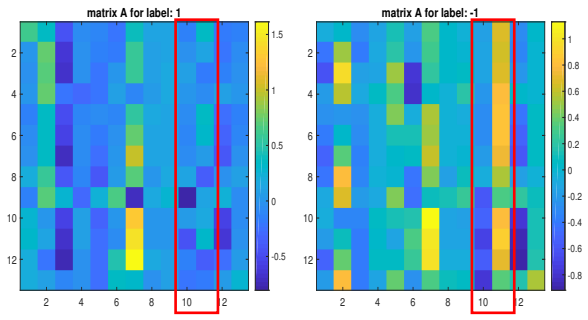
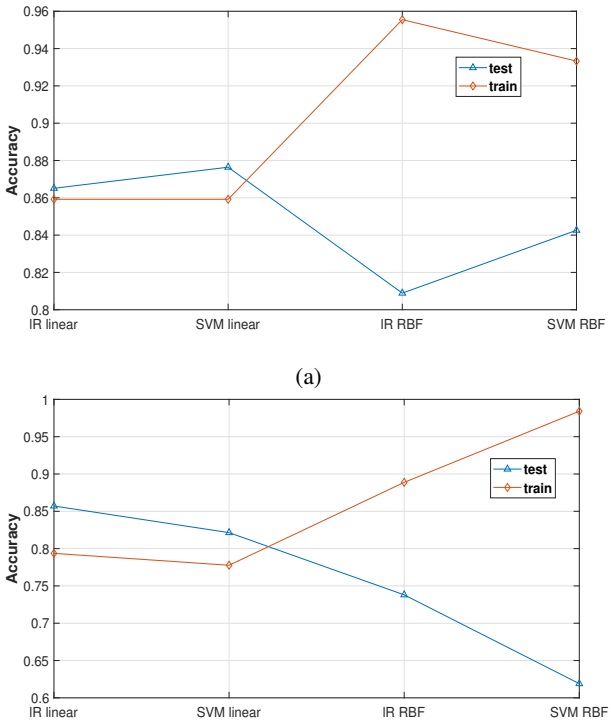


Fig. 12: Estimated A matrix of size 13×13 for the dataset-II with marked columns corresponding to the sensor index 10 and 11 used for classification.



(a)

(b)

Fig. 13: Testing and training accuracies for various classifiers arranged in the order of classification model complexity from left to right. The estimated accuracies for dataset-I and dataset-II are shown in (a) and (b) respectively.

can see the classic machine learning divergence curve for both the datasets. The accuracy for training data increases when increasing the classification model complexity while it reduces for the testing data. This is intuitive because a complex classification model would try to better classify the training data. But the performance of the test data would reduce due to overfitting upon using the complex models. We have very few training examples to build the classifier and hence such trend is expected. The performance of the classifiers for both the datasets are fairly high which reflects the strength of the

estimated features. We can see a 87.6% test accuracy for dataset-I and 85.7% for dataset-II. While these accuracies depend a lot on the cross-validation numbers and other factors like choice of classifier which can be better tuned to get higher numbers.

For both the datasets we have seen that the proposed methodology efficiently extracts the features which can be used to differentiate the imagined motor movements. By implicitly removing the effects of the unwanted stimuli, the coefficients of the coupling matrix A are shown to be sufficient for discriminating relation between various EEG signals which are indicative of the motor movements. The testing accuracies are high which indicate the good quality of the extracted features.

IV. Conclusion

We have revisited the EEG-based noninvasive brain interfaces feature extraction and translation from a cyber-physical systems' lens. Specifically, we leveraged spatiotemporal fractional-order models that cope with the unknown inputs. The fractional-order models provide us the dynamic coupling changes that rule the EEG data collected from the different EEG sensors, and the fractional-order exponents capture the long-term memory of the process. Subsequently, unknown stimuli can be determined as the external input that least conforms with the fractional-order model. By doing so, we can filter-out from the brain EEG signals the unknown inputs, that might be originated in deeper brain structures as the result of the structural connectivity of the brain that crisscrosses different regions, or due to artifacts originated in the muscles (e.g., eye blinking or head movement). As a consequence, the filtered signal does not need to annihilate an entire band in the frequency domain, thus keeping information about some frequency regions of the signal that would be otherwise lost.

We have shown how the different features obtained from the proposed model can be used towards rethinking the EEG-based noninvasive interfaces. In particular, two datasets used in BCI competitions were used to validate the performance of the methodology introduced in this manuscript, which is compatible with some of the state-of-the-art performances while requiring a relatively small number of training points. We believe that the proposed methodology can be used within the context of different neurophysiological processes and corresponding sensing technologies. Future research will focus on leveraging additional information from the unknown inputs retrieved to anticipate specific artifacts and enable the deployment of neuro-wearables in the context of real-life scenarios. Furthermore, the presented methodology can be used as an exploratory tool by neuroscientists and physicians, by testing input and output responses and tracking their impact in the unknown inputs retrieved by the algorithm proposed; in other words, one will be able to systematically identify the origin and dynamics of stimulus across space and time. Finally, it would be interesting to explore the proposed approach in the closed-loop context, where the present models would benefit from control-like strategies to enhance the brain towards certain tasks or attenuate side effects of certain neurodegenerative diseases or disorders.

References

[1] E. Strickland, "5 Neuroscience Experts Weigh in on Elon Musk's Mysterious 'Neural Lace' Company," *IEEE Spectrum*, Apr 2017. [Online]. Available: https://spectrum.ieee.org/the-human-os/biomedical/devices/5-neuroscience-experts-weigh-in-on-elon-musks-mysterious-neural-lace-company

[2] ——, "Facebook Announces "Typing-by-Brain" Project," *IEEE Spectrum*, Apr 2017. [Online]. Available: https://spectrum.ieee.org/the-human-os/biomedical/bionics/facebook-announces-typing-by-brain-project

[3] ——, "Silicon Valleys Latest Craze: Brain Tech," *IEEE Spectrum*, Jun 2017. [Online]. Available: https://spectrum.ieee.org/biomedical/devices/silicon-valleys-latest-craze-brain-tech

[4] J. Wolpaw and E. W. Wolpaw, *Brain-computer interfaces: principles and practice*. OUP USA, 2012.

[5] J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, and A. Kübler, "An MEG-based brain–computer interface (BCI)," *Neuroimage*, vol. 36, no. 3, pp. 581–593, 2007.

[6] R. Sitaram, A. Caria, R. Veit, T. Gaber, G. Rota, A. Kuebler, and N. Birbaumer, "Fmri brain-computer interface: a tool for neuroscientific research and treatment," *Computational intelligence and neuroscience*, 2007.

[7] N. Weiskopf, K. Mathiak, S. W. Bock, F. Scharnowski, R. Veit, W. Grodd, R. Goebel, and N. Birbaumer, "Principles of a brain-computer interface (bci) based on real-time functional magnetic resonance imaging (fmri)," *IEEE transactions on biomedical engineering*, vol. 51, no. 6, pp. 966–970, 2004.

[8] S. M. Coyle, T. E. Ward, and C. M. Markham, "Brain–computer interface using a simplified functional near-infrared spectroscopy system," *Journal of neural engineering*, vol. 4, no. 3, p. 219, 2007.

[9] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[10] F. Lotte, "A tutorial on eeg signal-processing techniques for mental-state recognition in brain–computer interfaces," in *Guide to Brain-Computer Music Interfacing*. Springer, 2014, pp. 133–161.

[11] N. Brodu, F. Lotte, and A. Lécuyer, "Exploring two novel features for eeg-based brain–computer interfaces: Multifractal cumulants and predictive complexity," *Neurocomputing*, vol. 79, pp. 87–94, 2012.

[12] G. Gupta, S. Pequito, and P. Bogdan, "Dealing with unknown unknowns: Identification and selection of minimal sensing for fractional dynamics with unknown inputs," *under review for the American Control Conference 2018*, 2018. [Online]. Available: https://drive.google.com/open?id=0B-qhdmwC5KqRNlA4cDZ6NG05N2s

[13] B. Blankertz, K. R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. R. Millan, M. Schroder, and N. Birbaumer, "The bci competition iii: validating alternative approaches to actual bci problems," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, June 2006.

[14] E. Başar, C. Başar-Eroglu, S. Karakaş, and M. Schürmann, "Gamma, alpha, delta, and theta oscillations govern cognitive processes," *International journal of psychophysiology*, vol. 39, no. 2, pp. 241–248, 2001.

[15] A. K. Engel, P. Fries, and W. Singer, "Dynamic predictions: oscillations and synchrony in top-down processing," *Nature reviews. Neuroscience*, vol. 2, no. 10, p. 704, 2001.

[16] G. Buzsaki, *Rhythms of the Brain*. Oxford University Press, 2006.

[17] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proceedings of the IEEE*, vol. 89, no. 7, pp. 1123–1134, 2001.

[18] T. McMillen, T. Williams, and P. Holmes, "Nonlinear muscles, passive viscoelasticity and body taper conspire to create neuromechanical phase lags in anguilliform swimmers," *PLoS computational biology*, vol. 4, no. 8, p. e1000157, 2008.

[19] Y. Kobayashi, H. Watanabe, T. Hoshi, K. Kawamura, and M. G. Fujie, "Viscoelastic and nonlinear liver modeling for needle insertion simulation," in *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery*. Springer, 2012, pp. 41–67.

[20] C. Wex, M. Fröhlich, K. Brandstädter, C. Bruns, and A. Stoll, "Experimental analysis of the mechanical behavior of the viscoelastic porcine pancreas and preliminary case study on the human pancreas," *Journal of the mechanical behavior of biomedical materials*, vol. 41, pp. 199–207, 2015.

[21] K. Wang, R. McCarter, J. Wright, J. Beverly, and R. Ramirez-Mitchell, "Viscoelasticity of the sarcomere matrix of skeletal muscles. the titin-myosin composite filament is a dual-stage molecular spring," *Biophysical Journal*, vol. 64, no. 4, pp. 1161–1177, 1993.

[22] T. M. Best, J. McElhaney, W. E. Garrett, and B. S. Myers, "Characterization of the passive responses of live skeletal muscle using the quasi-linear theory of viscoelasticity," *Journal of biomechanics*, vol. 27, no. 4, pp. 413–419, 1994.

[23] T. C. Doehring, A. D. Freed, E. O. Carew, and I. Vesely, "Fractional order viscoelasticity of the aortic valve cusp: an alternative to quasilinear viscoelasticity," *Journal of biomechanical engineering*, vol. 127, no. 4, pp. 700–708, 2005.

[24] E. Macé, I. Cohen, G. Montaldo, R. Miles, M. Fink, and M. Tanter, "In vivo mapping of brain elasticity in small animals using shear wave imaging," *IEEE transactions on medical imaging*, vol. 30, no. 3, pp. 550–558, 2011.

[25] S. Nicolle, L. Noguer, and J.-F. Palierne, "Shear mechanical properties of the spleen: experiment and analytical modelling," *Journal of the mechanical behavior of biomedical materials*, vol. 9, pp. 130–136, 2012.

[26] N. Grahovac and M. Žigić, "Modelling of the hamstring muscle group by use of fractional derivatives," *Computers & Mathematics with Applications*, vol. 59, no. 5, pp. 1695–1700, 2010.

[27] V. E. Tarasov, *Fractional dynamics: applications of fractional calculus to dynamics of particles, fields and media*. Springer Science & Business Media, 2011.

[28] P. Bogdan, B. M. Deasy, B. Gharaibeh, T. Roehrs, and R. Marculescu, "Heterogeneous structure of stem cells dynamics: statistical models and quantitative predictions," *Scientific reports*, vol. 4, 2014.

[29] M. Ghorbani and P. Bogdan, "A cyber-physical system approach to artificial pancreas design," in *Proceedings of the ninth IEEE/ACM/IFIP international conference on hardware/software codesign and system synthesis*. IEEE Press, 2013, p. 17.

[30] Y. Xue, S. Rodriguez, and P. Bogdan, "A spatio-temporal fractal model for a cps approach to brain-machine-body interfaces," in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2016*. IEEE, 2016, pp. 642–647.

[31] Y. Xue, S. Pequito, J. R. Coelho, P. Bogdan, and G. J. Pappas, "Minimum number of sensors to ensure observability of physiological systems: A case study," in *Communication, Control, and Computing (Allerton), 2016 54th Annual Allerton Conference on*. IEEE, 2016, pp. 1181–1188.

[32] Y. Xue and P. Bogdan, "Constructing compact causal mathematical models for complex dynamics," in *Proceedings of the 8th International Conference on Cyber-Physical Systems*. ACM, 2017, pp. 97–107.

[33] A. Dzielinski and D. Sierociuk, "Adaptive feedback control of fractional order discrete state-space systems," in *CIMCA-IAWTIC*, 2005.

[34] P. Flandrin, "Wavelet analysis and synthesis of fractional brownian motion," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 910–917, March 1992.

[35] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. John Wiley & Sons, New York, 1996.

[36] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[37] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Mller, "Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun 2004.

[38] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in braincomputer interfaces based on electrical brain signals," *Journal of Neural Engineering*, vol. 4, no. 2, p. R32, 2007.

[39] N. Brodu, F. Lotte, and A. Lcuyer, "Comparative study of band-power extraction techniques for motor imagery classification," in *2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, April 2011, pp. 1–6.

[40] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle, "Comparative analysis of spectral approaches to feature extraction for eeg-based motor imagery classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 16, no. 4, pp. 317–326, Aug 2008.

Decoding the Principles of Emergence and Resiliency in Biological Collective Systems - A Multi-scale Approach

Paul Bogdan and James Boedicker

University of Southern California

{pbogdan, boedicke}@usc.edu

DARPA Deep Purposeful Learning (Deep Purple) and Fundamental of Complex Collectives (FunCC) Kickoff Meeting

November 14-17th, 2017

USC

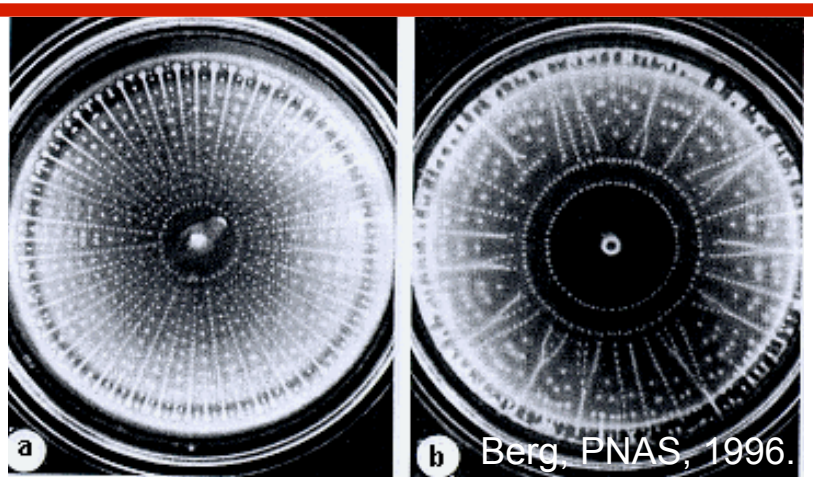
DARPA

Project Outline

- **Pattern formation diversity in wild microbial societies**
  - Experimental and mathematical analysis methodology
  - Skeleton for a theory of complex collectives
- **Noise and its consequences on collective behavior**
  - Computational quantification of pattern formation sensitivity to noise
  - Experimental validation
- **Impact of single-cell heterogeneity on pattern formation**
  - Experimental quantification of single cell behavior within cellular collectives
  - Mathematical quantification of information transfer in microbial societies
  - Free-energy landscape description of collective biological systems
- **Robustness of pattern formation**
  - Predict the outcome of two systems of pattern forming cells
  - Implications for a theory of complex collectives

Pattern Formation Diversity in Microbial Society

- **Microscopic computation / learning, storage (multiscale retainment of information) and communication lead to complex macroscopic pattern formation**



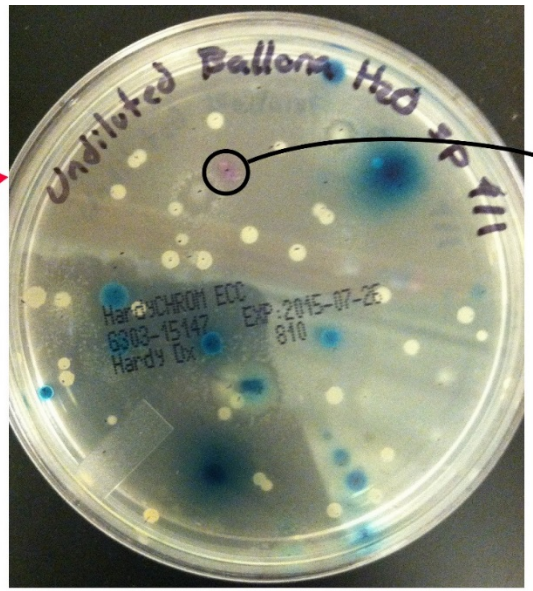
  - Single cell computation: environmental sensing + internal processing for decision making (expressing genes, moving faster/slower, compete or collaborate)
  - Memory storage: at single cell and at population level
  - Communication: convey information for self-organization and emergence
  - E.coli form elaborate patterns of rings and spots on a soft agar plate
- **Factors influencing pattern formation**
  - Chemotaxis, amino acid-mediated signaling (communication protocols)
  - Nutrient depletion (computational resources / constraints)

How to classify the pattern diversity formed by natural coliform isolates?

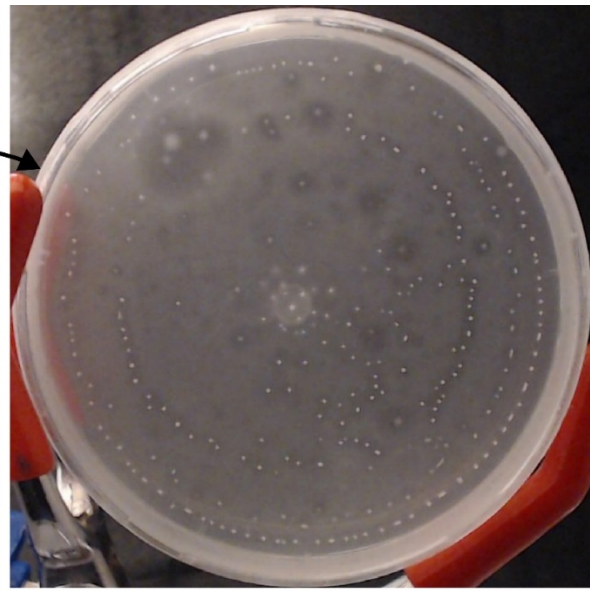
Diversity of Collective Behaviors in the Wild

- **Many wild bacteria were isolated from LA fresh water samples and screened for collective pattern formation.**

Collect environmental sample

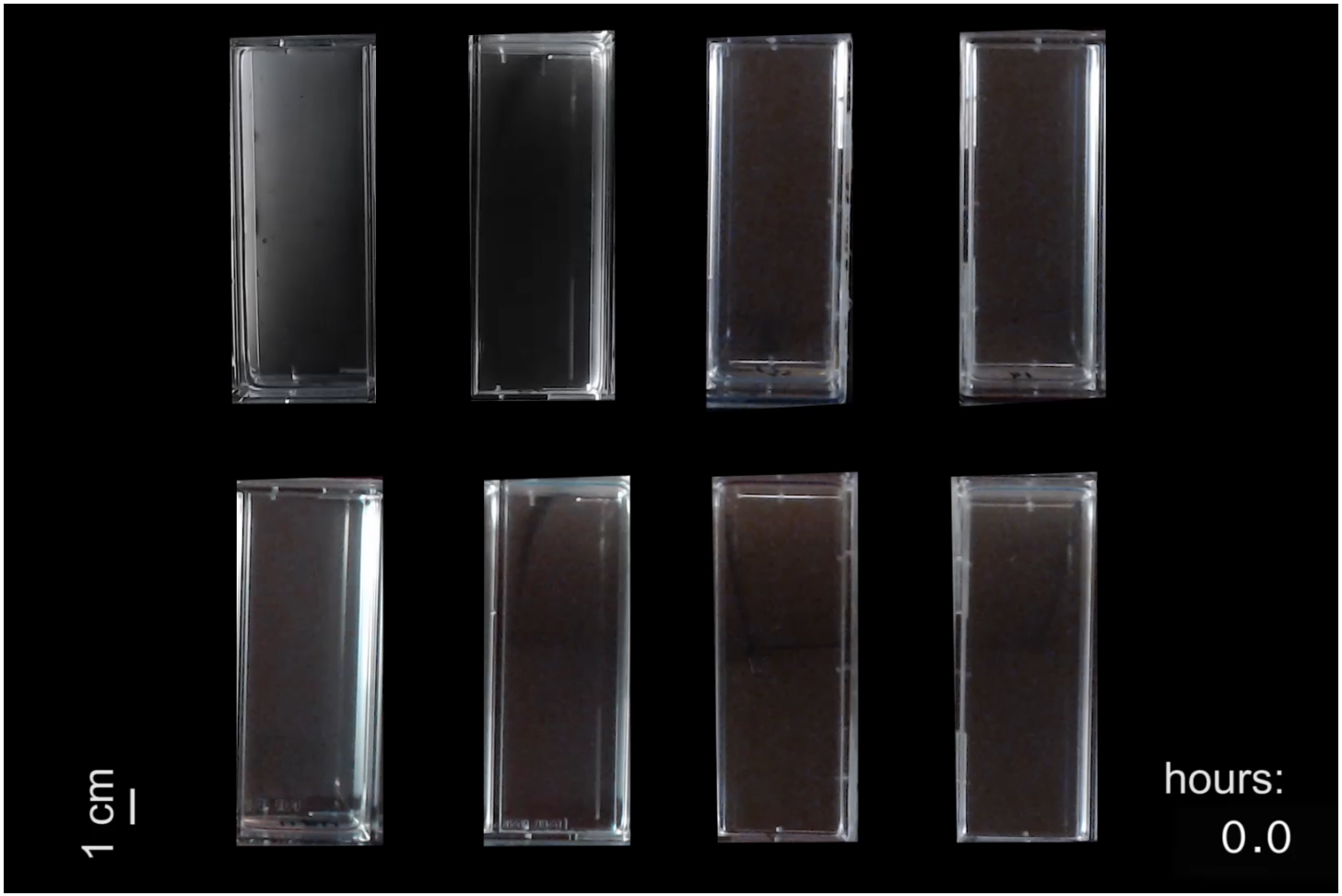


Isolate *E. coli* cell on selection plate



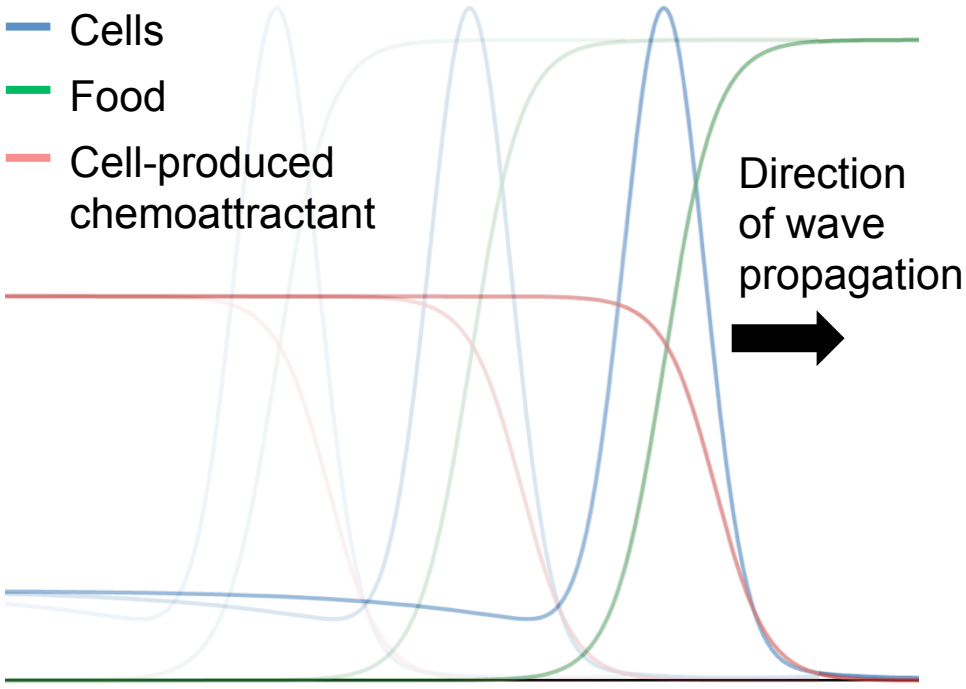
Test for ability to form patterns of spots and rings

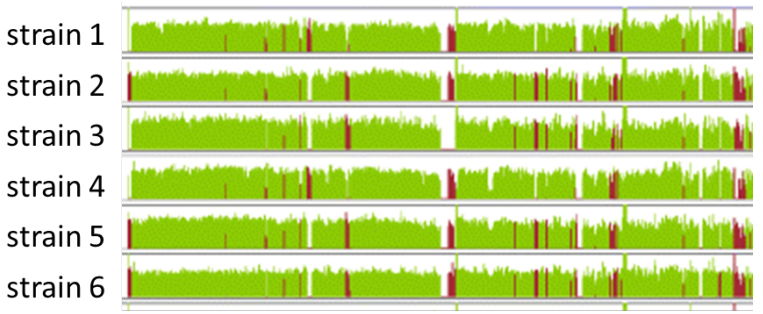
Isolates form similar but distinct dynamic patterns

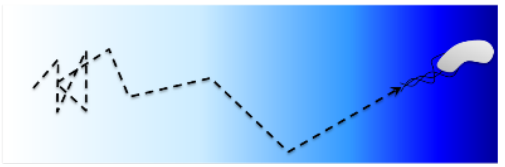


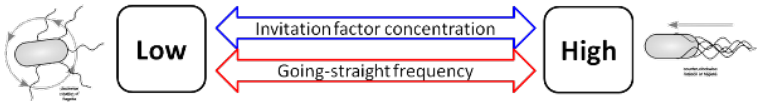
What differences at the genomic and microscopic level generate variability in emergent behavior?

- **The band forms as a result of chemotaxis and cell-cell signaling.**

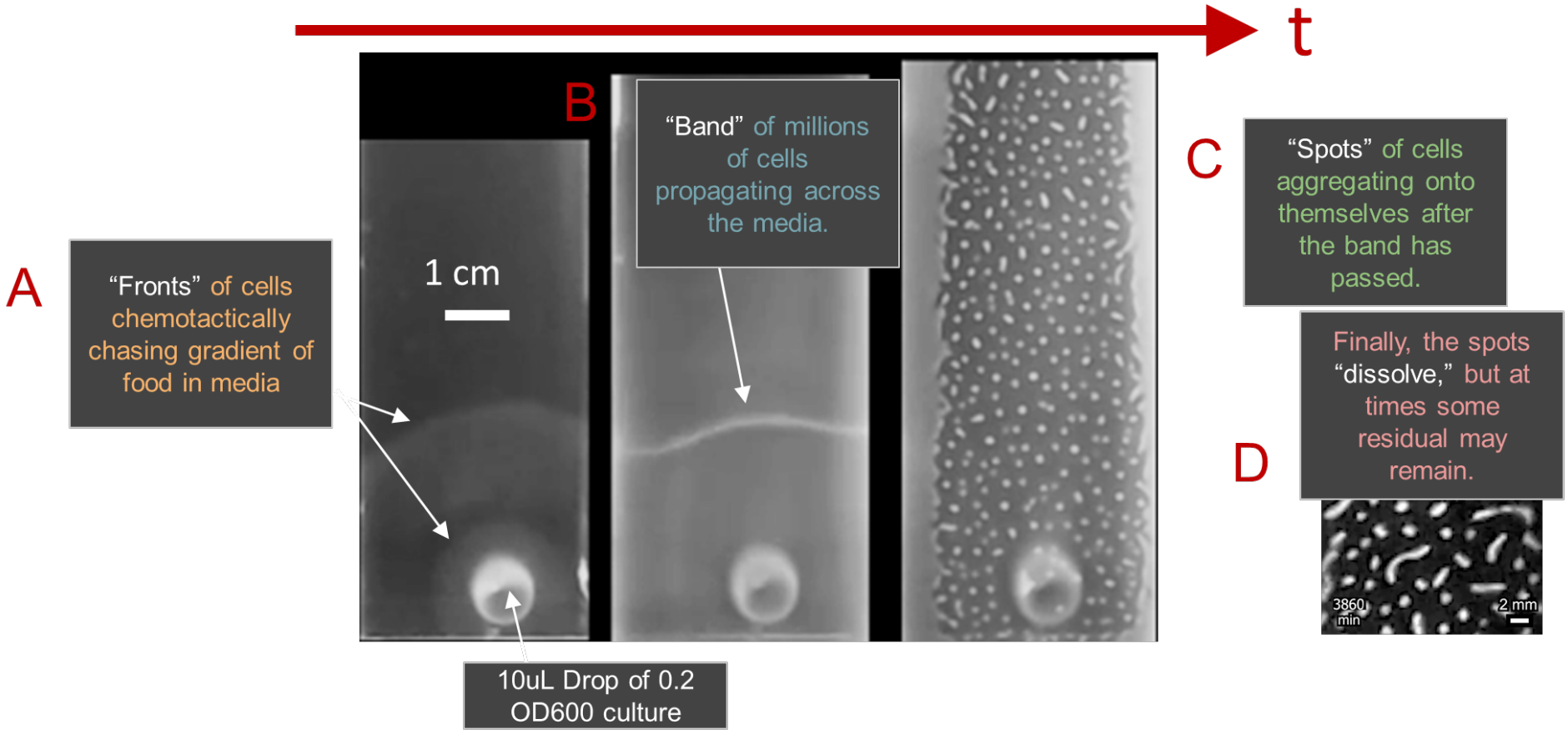




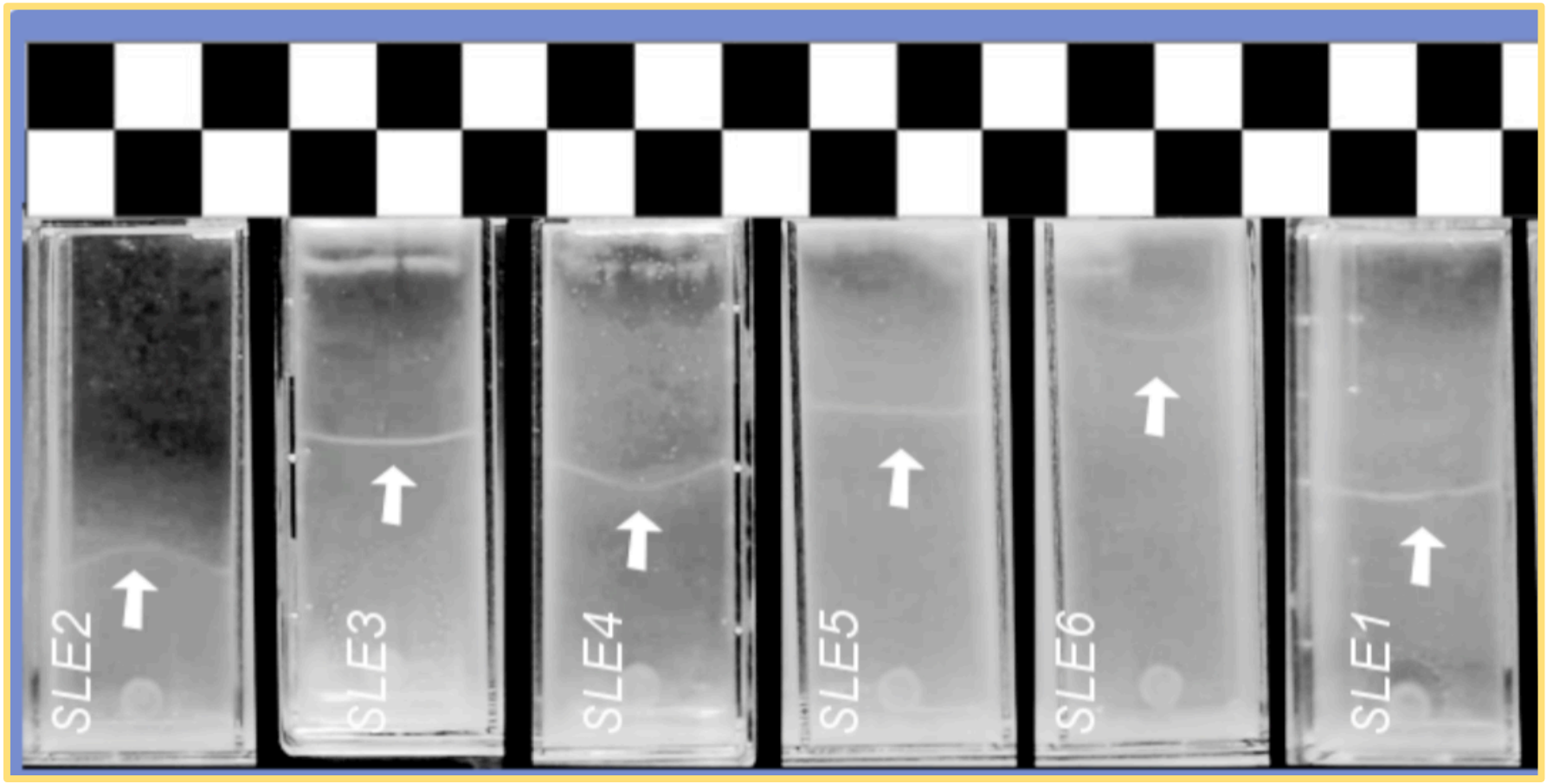




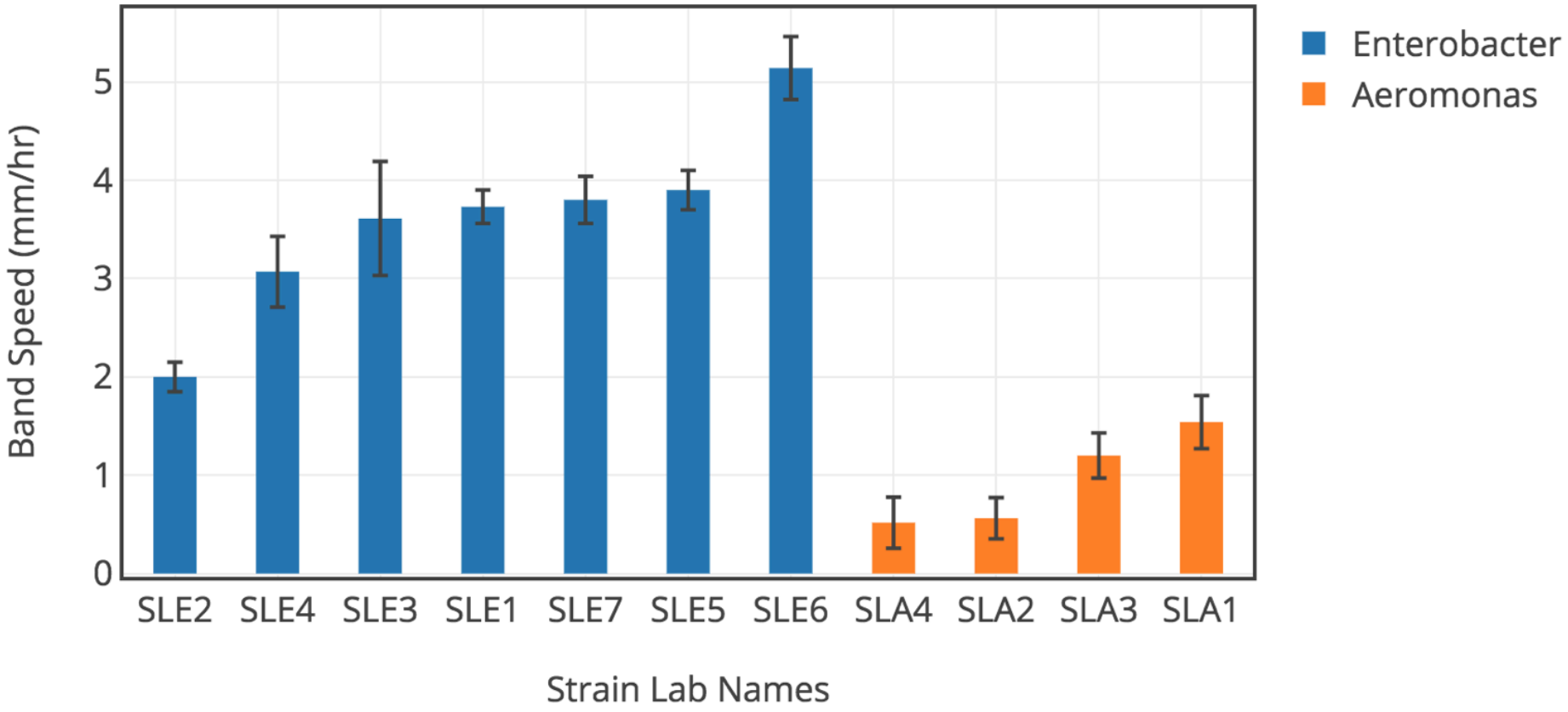
Multiple stages of pattern formation



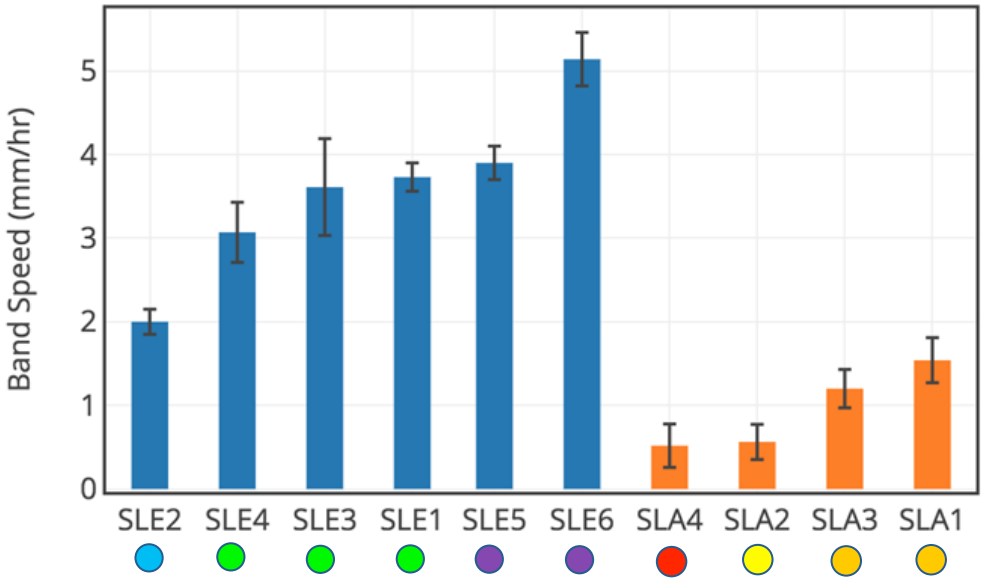
Microbial “races” revealed variability of the band speed



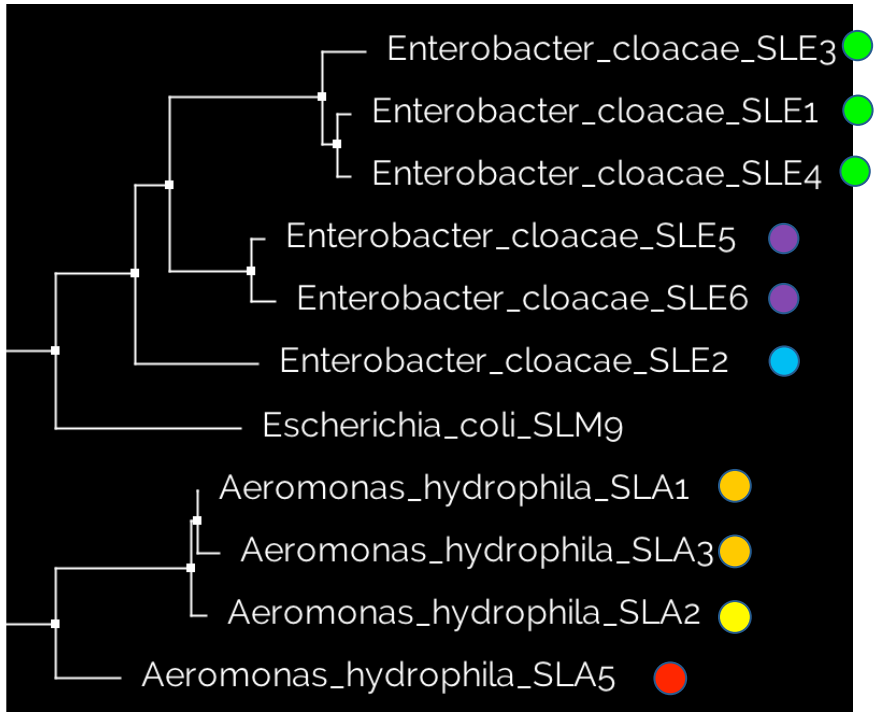
The velocity of the swarm ring was variable, even within closely related strains



Strain relatedness correlates with similarity in band speed

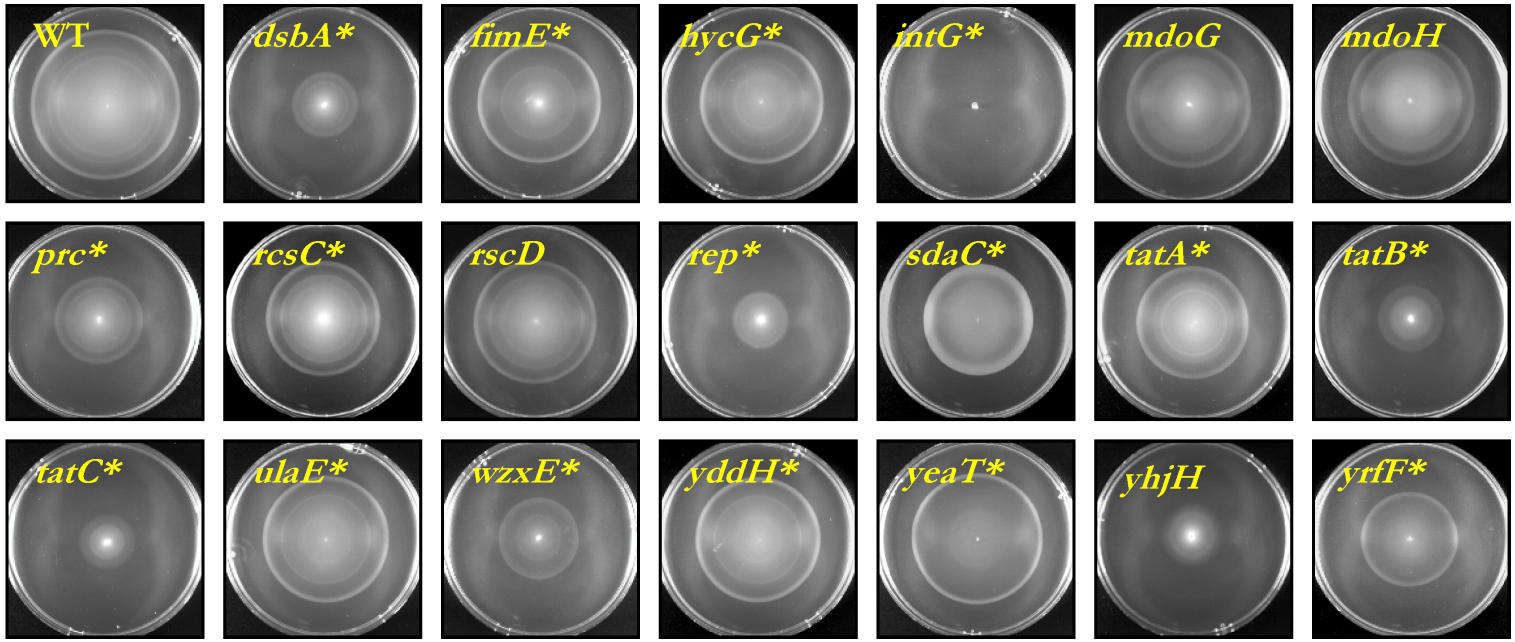


16S rRNA phylogenic tree



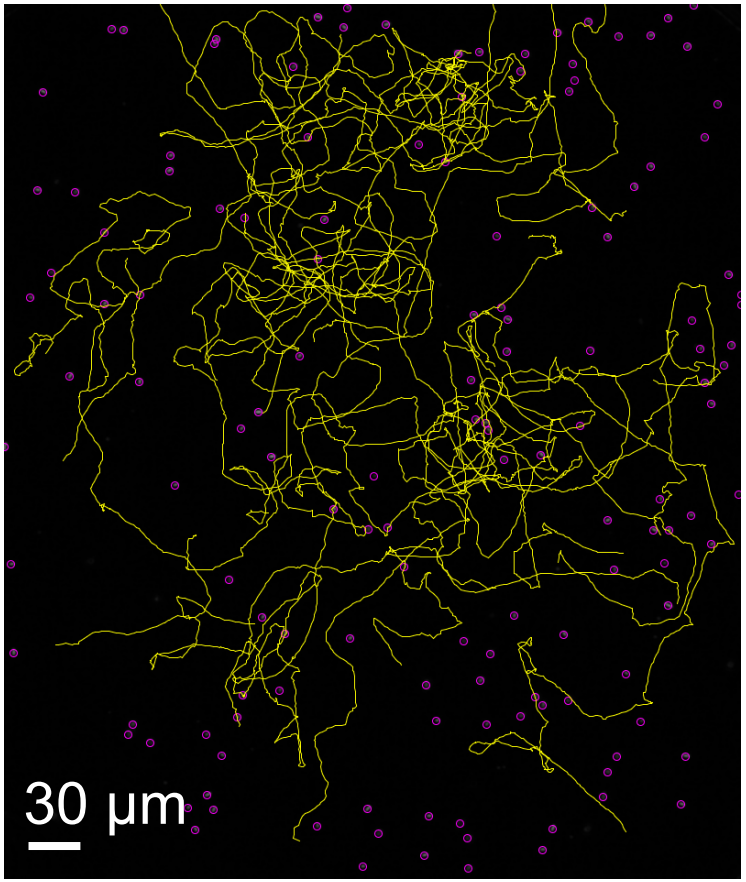
Genomic diversity modulates pattern formation

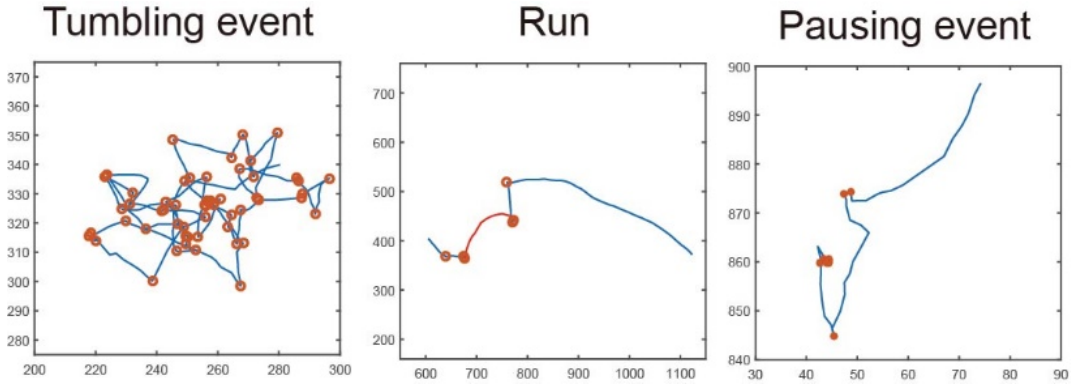
- **Multiple genes contribute to patterning, even those not obviously associated with motility (*).**
- **We have sequenced the genomes of 25 isolates to identify genomic changes that fine tune collective behavior.**
- **Thus far, all known pattern formation genes are present in each strain, although gene and promoter sequences are variable.**



HS Girgis, Y Liu, WS Ryu, and Saeed Tavazoie, A Comprehensive Genetic Characterization of Bacterial Motility, PLoS Genetics, 2007.

Measure the single-cell motility characteristics of each strain

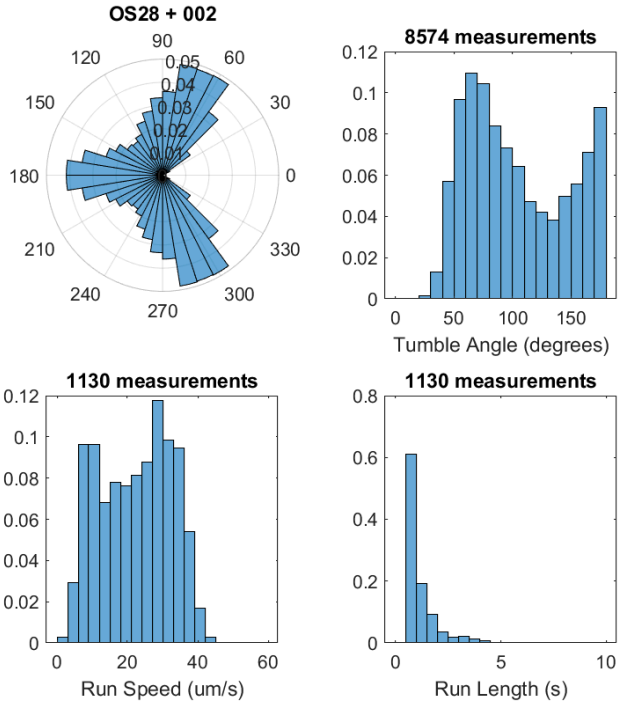


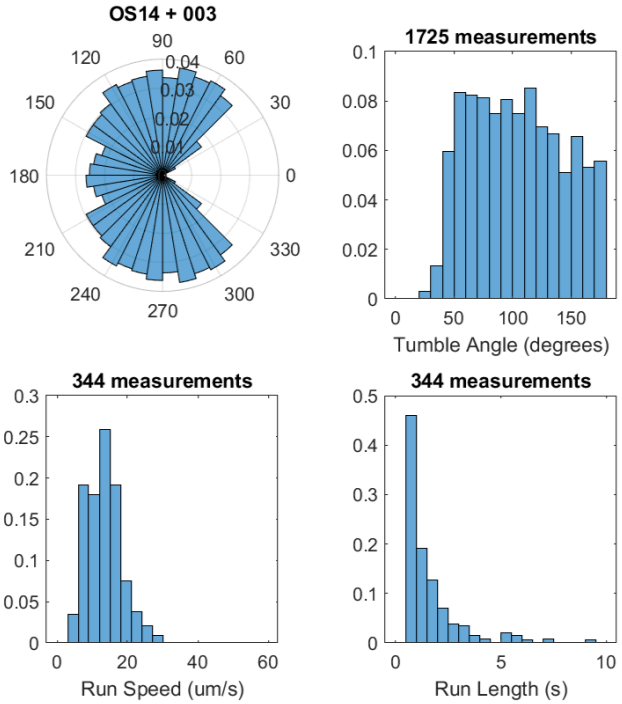


- **At the microscopic level, how do the swimming trajectories of each strain and species differ?**
- **Will differences in the behavior of individual agents predict the variability of the emergent patterns formed at larger scales?**

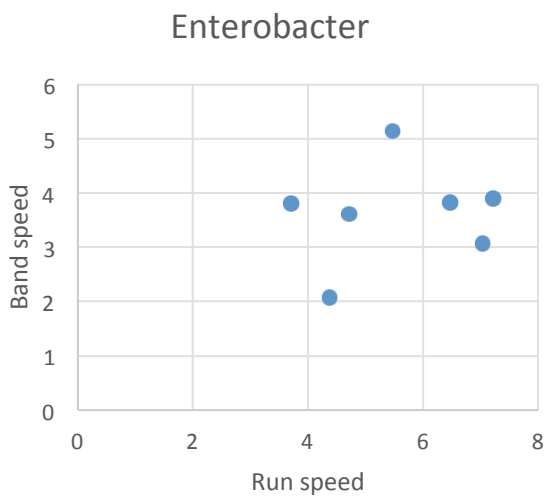
The two species that form bands have very different single-cell swimming behavior.

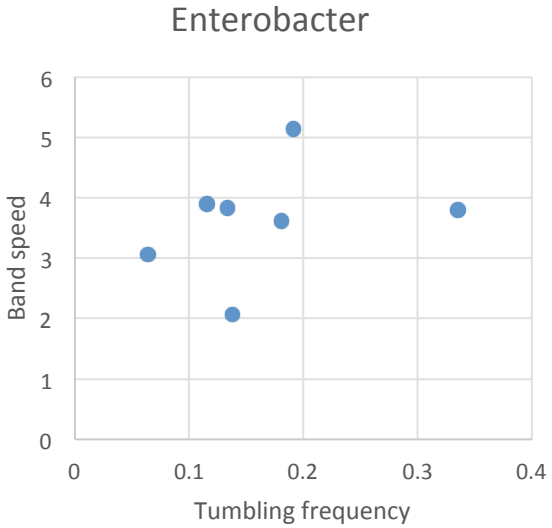
Aeromonas hydrophila SLA2	*Enterobacter cloacae SLE2*

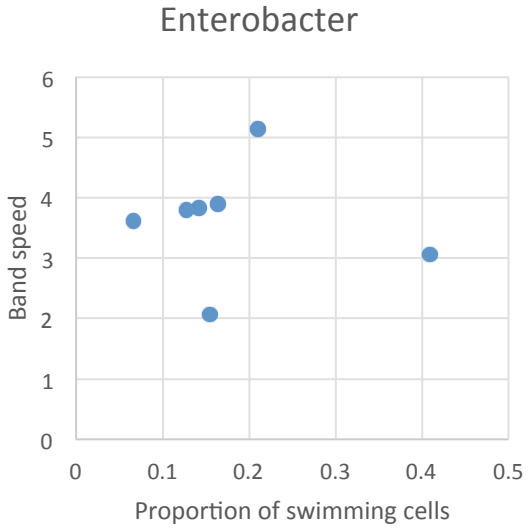


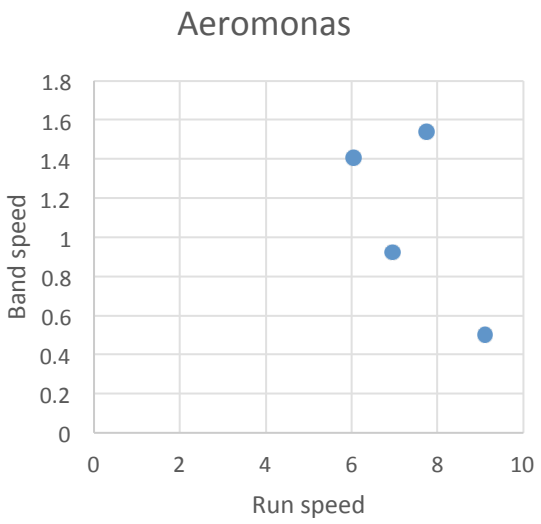


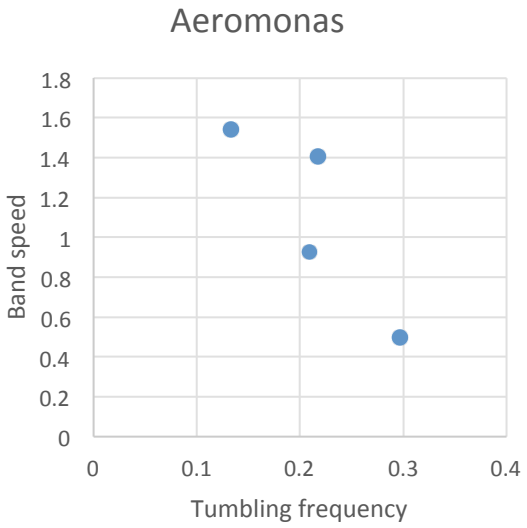
No strong correlations between the band speed and single-cell swimming characteristics

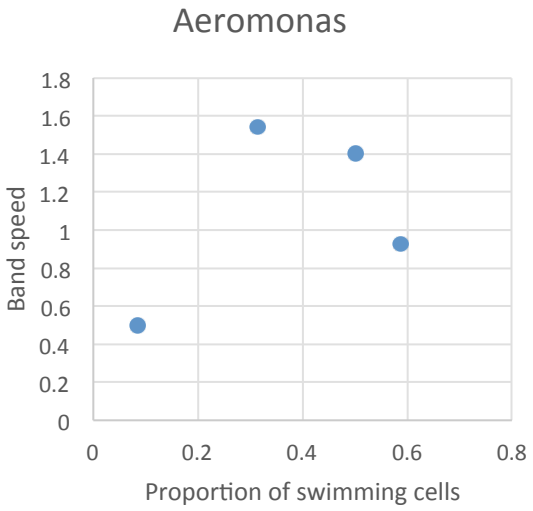












No strong correlations between the band speed and single-cell swimming characteristics

Describe the cells moving in 1-D:

$$\frac{\partial r}{\partial t} = -v\frac{\partial r}{\partial t} - f_{rl}r + f_{lr}l,$$

$$\frac{\partial l}{\partial t} = v\frac{\partial l}{\partial t} + f_{rl}r - f_{lr}l$$

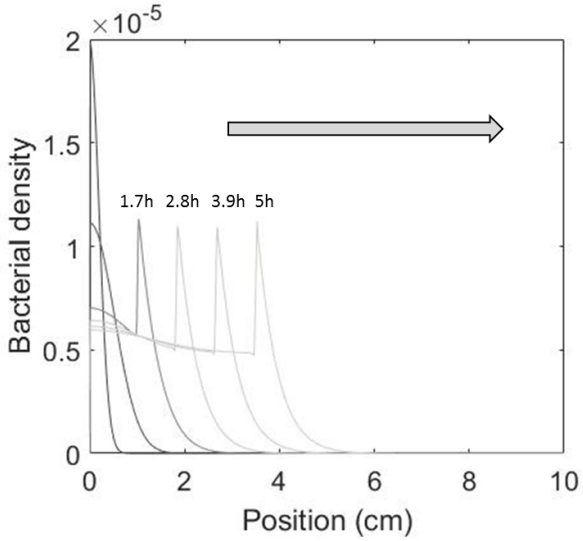
$r(x,t)$	Right-moving bacterial density
$l(x,t)$	Left-moving bacterial density
v	Constant swimming speed
f_{rl}	Frequency of the reversal from right to left
f_{lr}	Frequency of the reversal from left to right

Concentrations of nutrient and chemoattractant:

$$\frac{\partial N}{\partial t} = D_N\frac{\partial^2 N}{\partial x^2} - \kappa_N b,$$

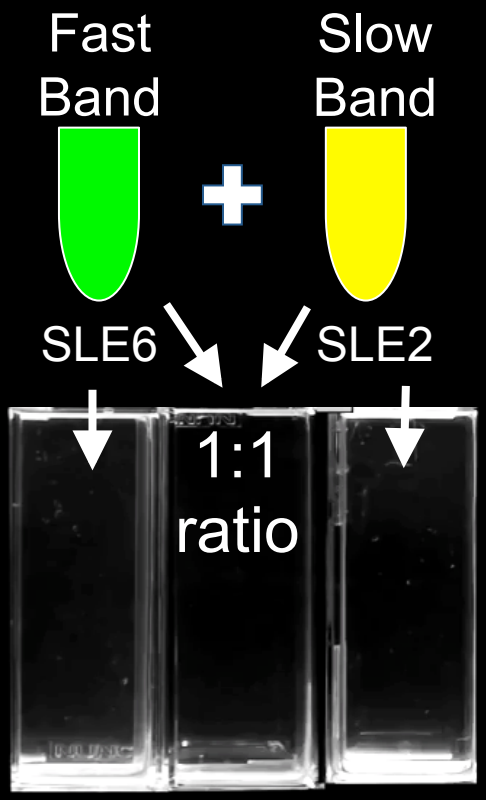
$$\frac{\partial S}{\partial t} = D_S\frac{\partial^2 S}{\partial x^2} - \mu_S S + \kappa_s b$$

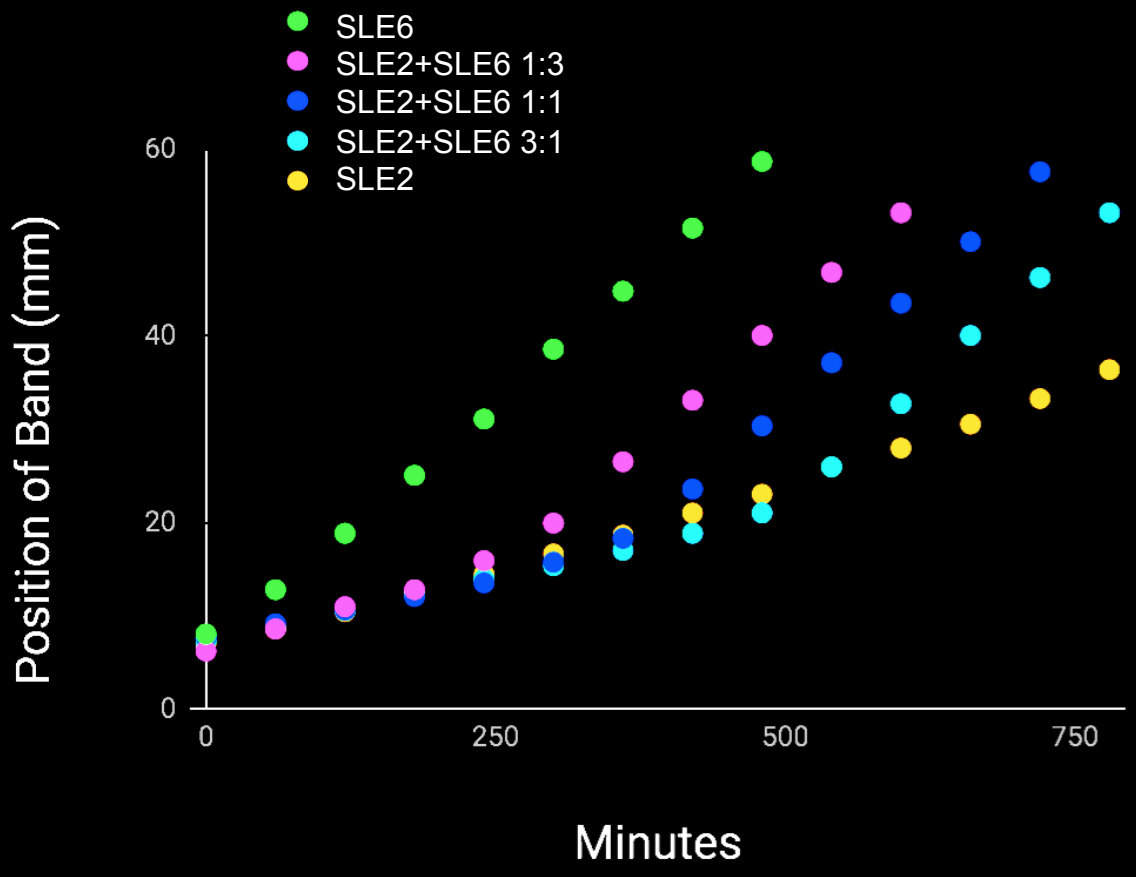
$\mathrm{N}(x,t)$	Nutrient concentration
$\mathrm{S}(x,t)$	Chemoattractant concentration
$b(x,t)$	Bacterial density
D_N	Nutrient diffusion coefficient
D_S	Chemoattractant diffusion coefficient
κ_N	Nutrient consumption rate
μ_S	Chemoattractant production rate
κ_s	Chemoattractant degradation rate



- **Band formation can be reproduced in continuum based models.**
- **Also pursuing agent based modeling that incorporates experimental measurements of single-cell behavior and its variability.**

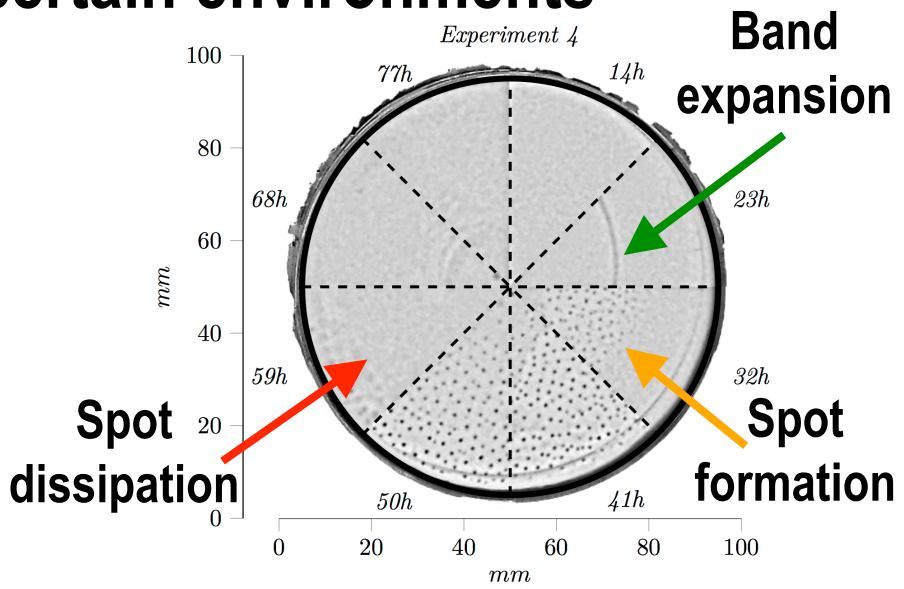
Hybrid patterns formed by combining multiple strains





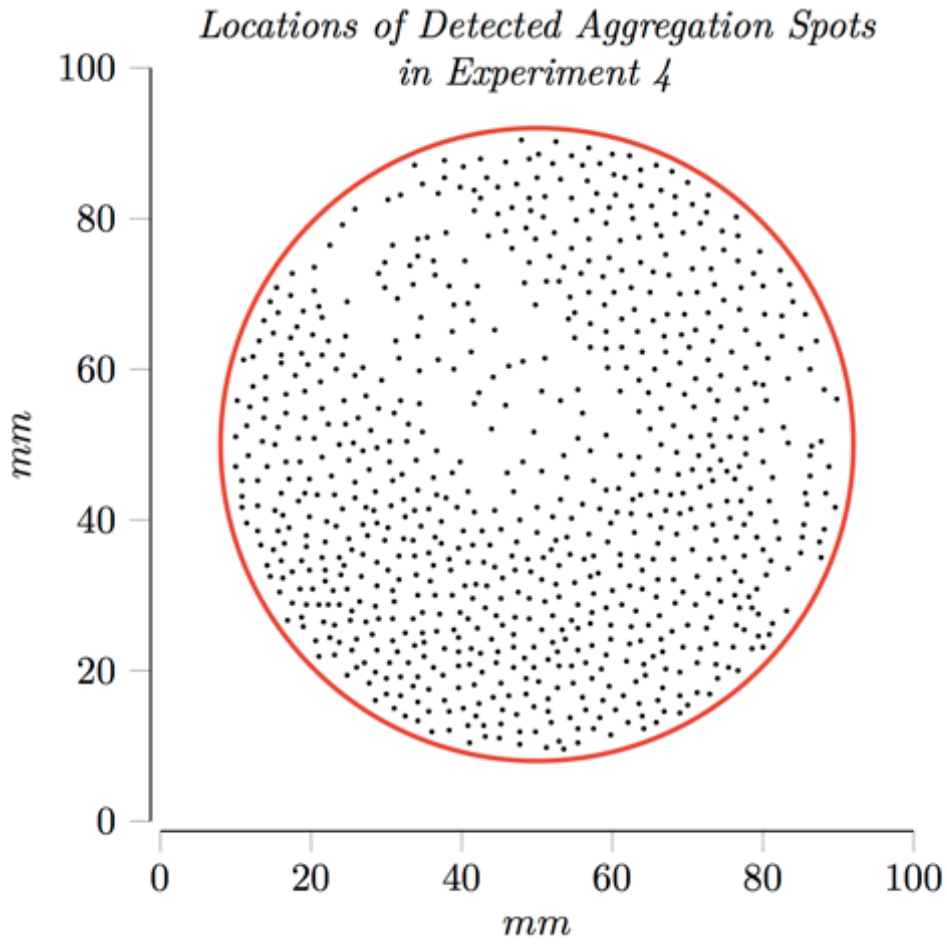
MC Exhibit Complex Spatio-Temporal Patterns

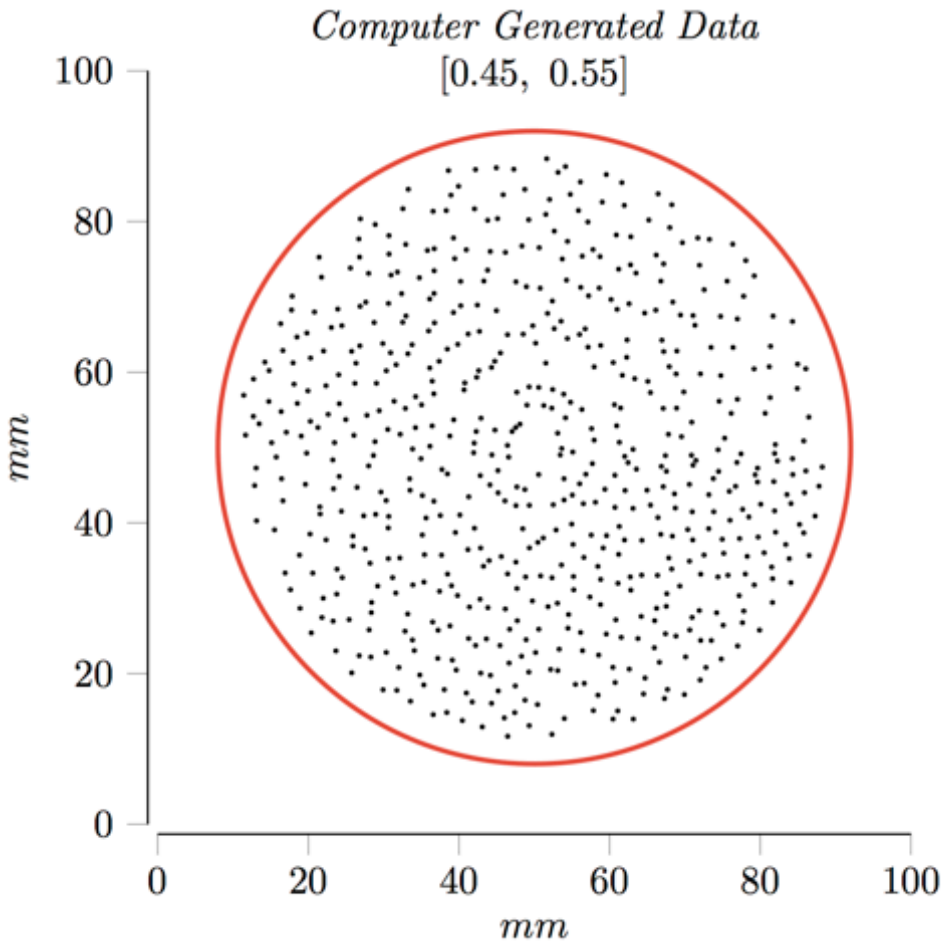
- **Bacteria aggregate patterns encode cooperative, competitive and adaptive interactions in uncertain environments**
  - Microscopic information is hard to obtain, scarce or inaccurate
  - Environment is continuously evolving
  - Difficult to track and separate various effects produced by interactions
- **Macro-scopic observations**
  - Band expansion (after 14h)
  - Spot formation (after 20-30h)
  - Spot dissipation (after 50h)
- **Complex signatures**
  - Between band expansion, spot aggregation and dissipation



Natural vs. Artificial Patterns

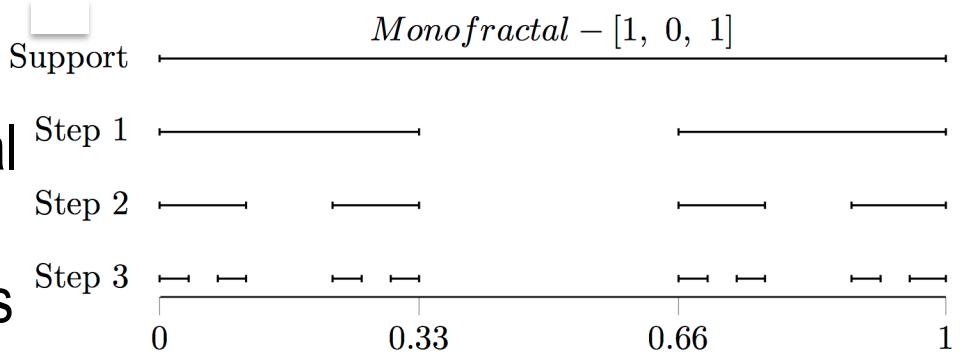
- **Hard to quantify what language collective microbial communities speak?**





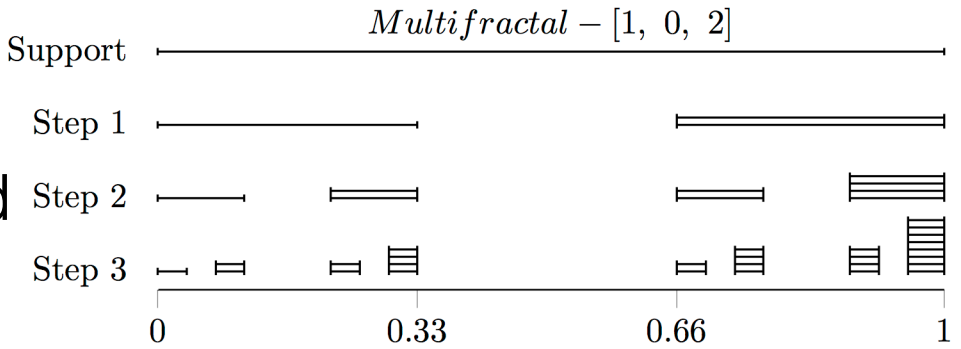
Mono-fractals vs. Multi-fractals (I)

- **Mono-fractals**
  - Patterns with self-similar statistical properties characterized by single fractal dimension across all scales



$$D(q) = \lim_{1/s \to 0} \frac{1}{1-q} \frac{\log\left(\sum_{i=1}^{n} p_i^q\right)}{\log(1/s)} = \frac{1}{1-q} \frac{\log\left(2^n \cdot \frac{1}{2^{qn}}\right)}{\log\left(1/3^n\right)} = \frac{\log(2)}{\log(3)}$$

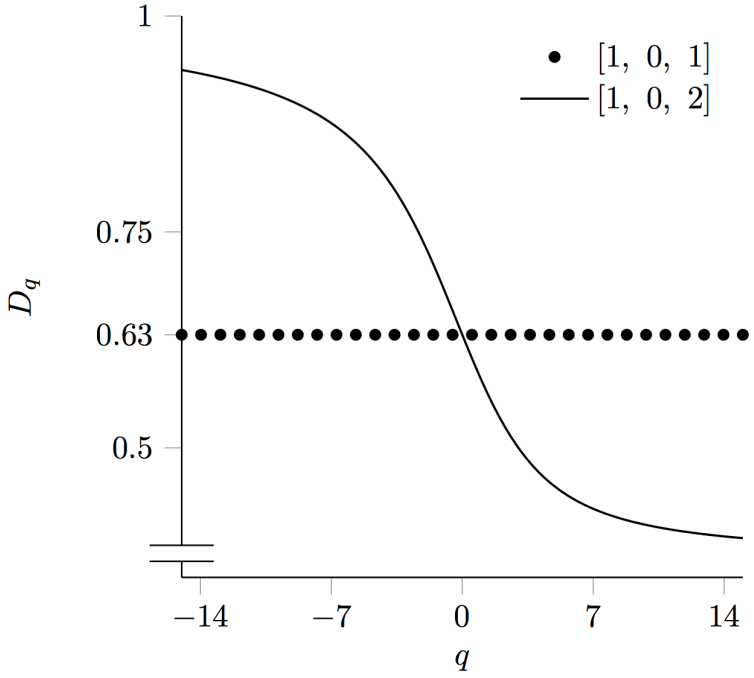
- **Multi-fractals**
  - Patterns with self-similar statistical properties characterized by different fractal dimensions across multiple scales

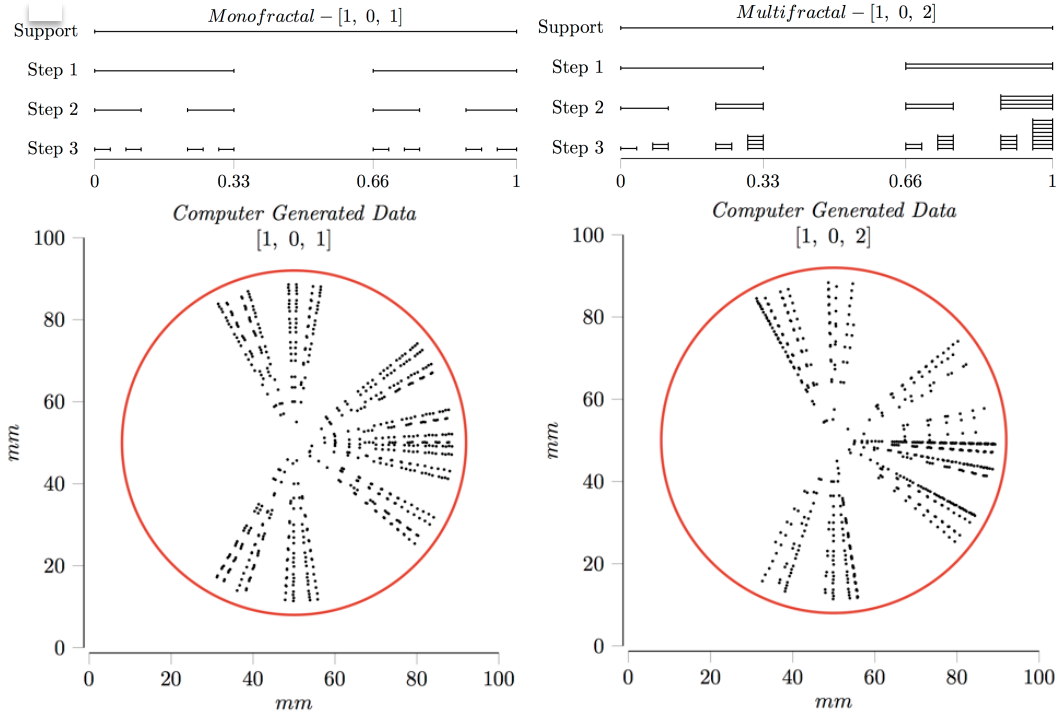


$$D(q) = \lim_{1/s \to 0} \frac{1}{1-q} \frac{\log\left(\sum_{i=1}^{n} p_i^q\right)}{\log(1/s)} = \frac{1}{1-q} \frac{\log\left(\left(\frac{1}{3}\right)^q + \left(\frac{2}{3}\right)^q\right)}{\log\left(1/3^n\right)}$$

Mono-fractals vs. Multi-fractals (II)

- **Mono-fractal Cantor set – inspired aggregation pattern [1 0 1]**
- **Multi-fractal Cantor set – inspired aggregation pattern [1 0 2]**

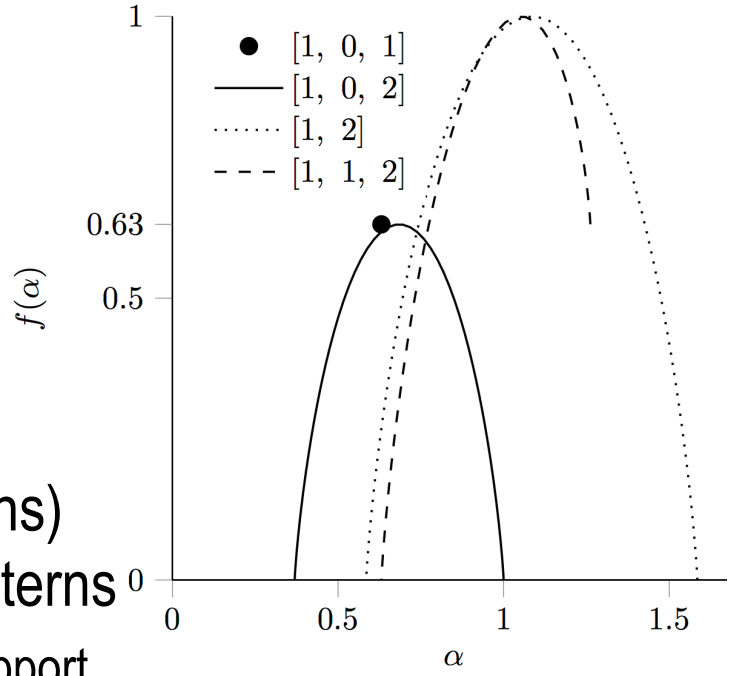


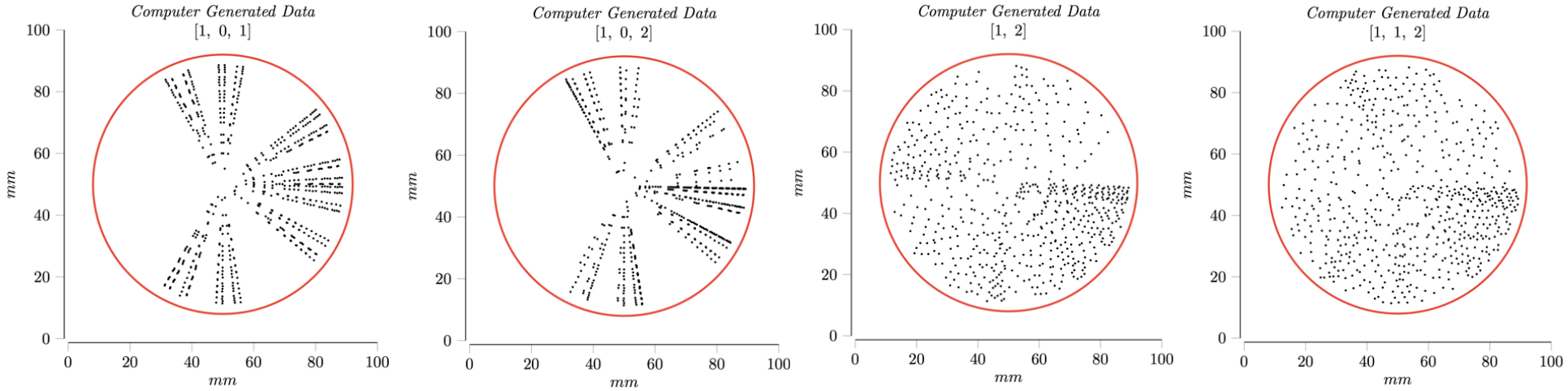


- **Generalized dimension D_q**
  - Detects deviations from perfect order
  - Quantifies the mean of the spatio-temporal distributions of the support (D_0), information dimension (D_1), correlation dimension (D_2), rare events ($D_{-\infty}$)

Mono-fractals vs. Multi-fractals (III)

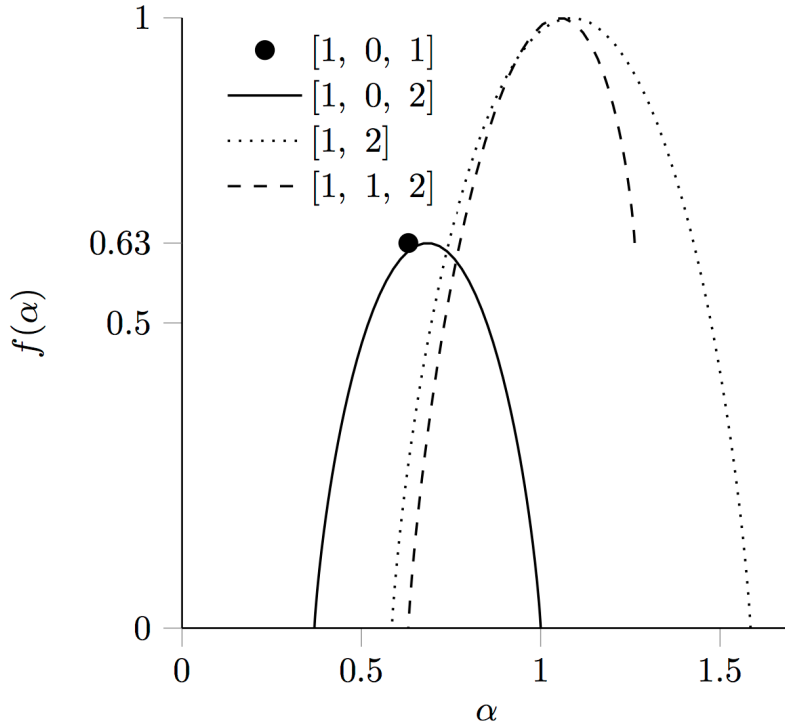
- **Multi-fractal spectrum (MFS) $f(\alpha)$**
  - Single point for mono-fractals
  - Wideness of multi-fractal spectrum measures the variability in properties across regions
  - α_{min} and α_{max} correspond to highest and smallest probabilities
- Patterns with zero probability (lacunar regions) have MFSs with lower peak than denser patterns
  - Peak shows how much the fractal covers the support

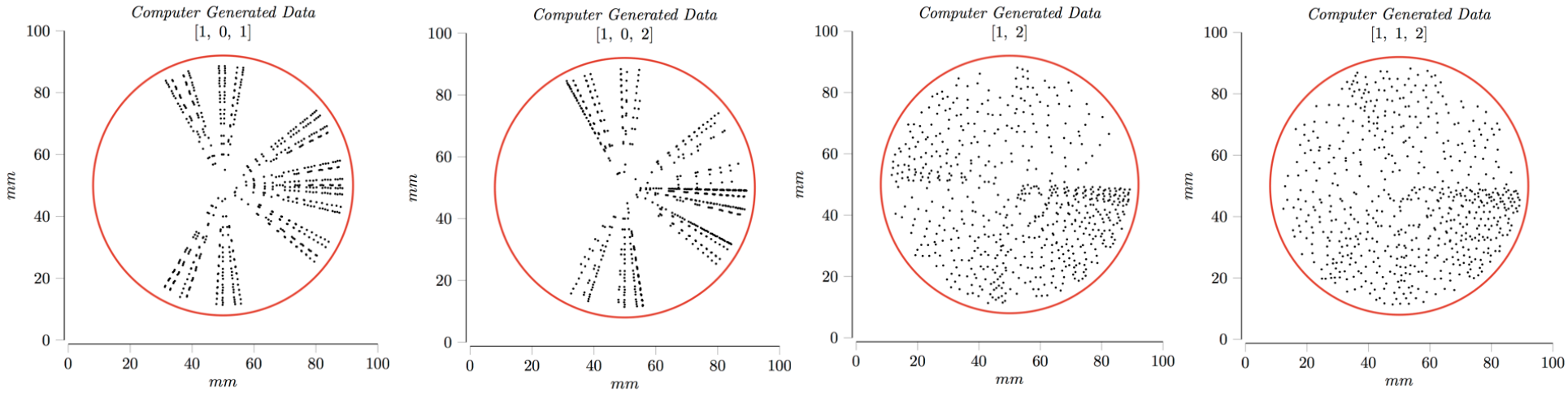




Mono-fractals vs. Multi-fractals (IV)

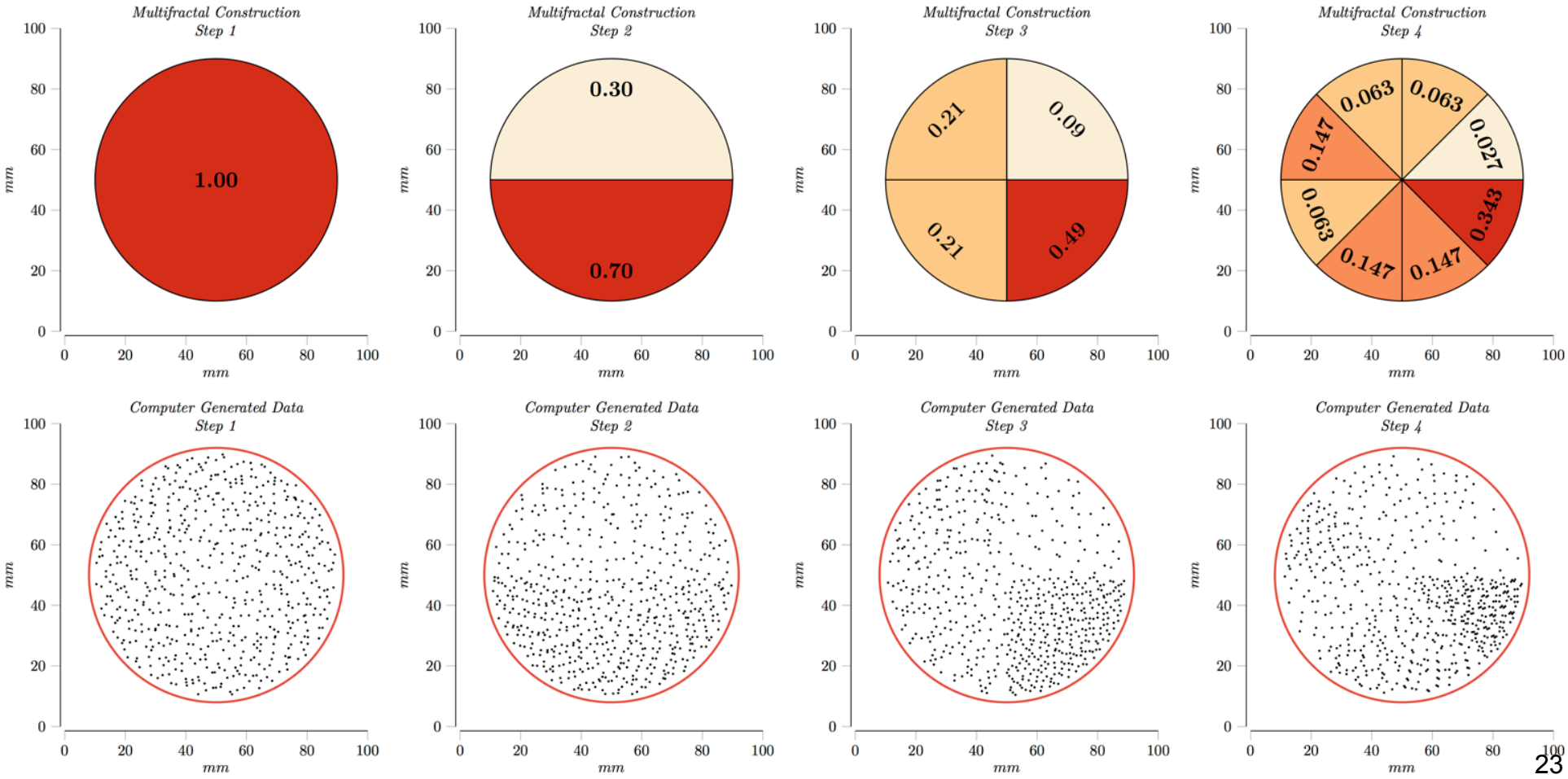
- **Multi-fractal spectrum $f(\alpha)$ (con't)**
  - Patterns with multiple scale dependent regions with similar properties have a multi-fractal spectrum $f(\alpha)$ greater than zero
    - [1 1 2] consists of repeated exact motifs hence its $f(\alpha_{max})$ will be greater than zero





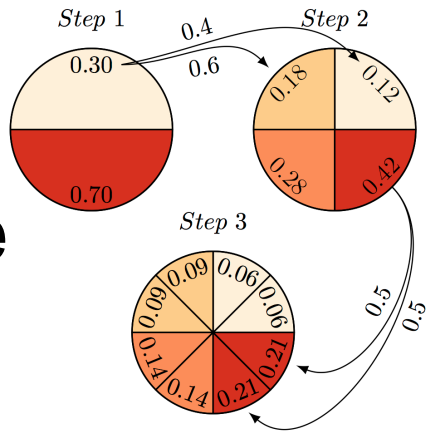
Artificial Multi-Fractal Patterns: Examples

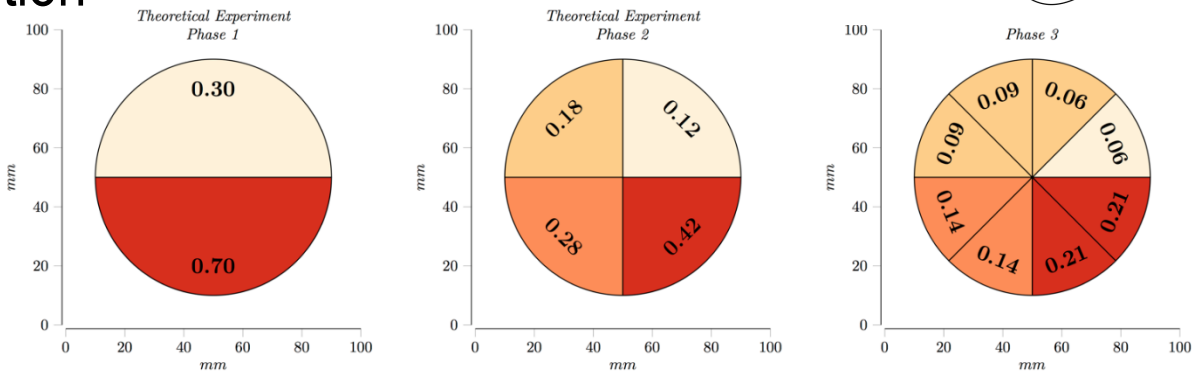
- **Assign 1 to the support interval**
- **Split in two each interval and follow weights (0.3 for left & 0.7 for right)**
- **Repeat step two indefinitely**

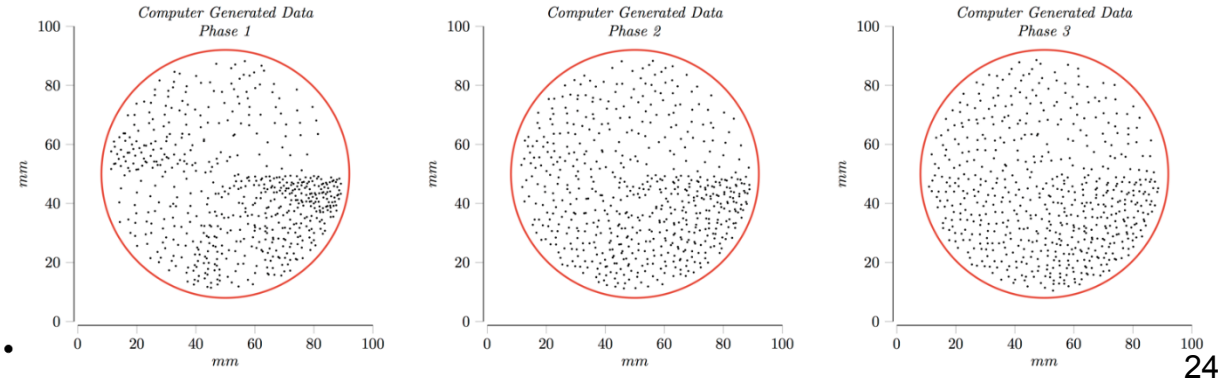


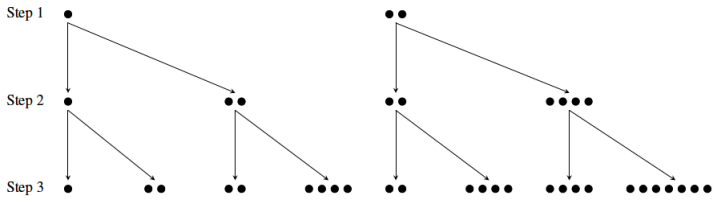
Implications

- **Controlling degree of order (multi-fractality) in an artificial pattern formation system**
- **Consider three pattern aggregation steps where**
  - Values in each sector represent the probability of an aggregation spot formation
  - Variation in probability of different sectors encodes the pattern heterogeneity (multi-fractality) or departure from mono-fractality







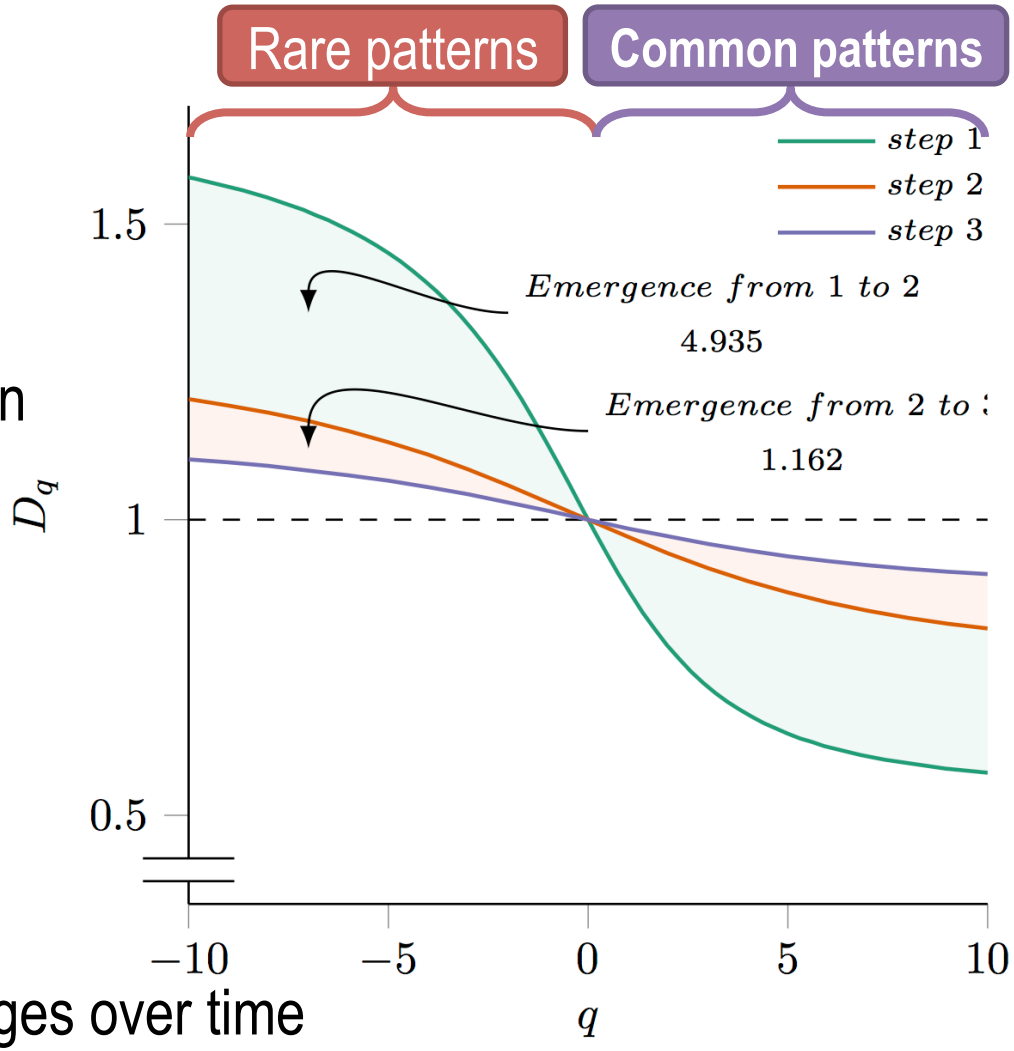


Emergence: Definition

- **Emergence**
  - Encodes a phase transition (system level property) and hysteresis phenomena
  - Quantifies amount of information generated in the whole by interactions

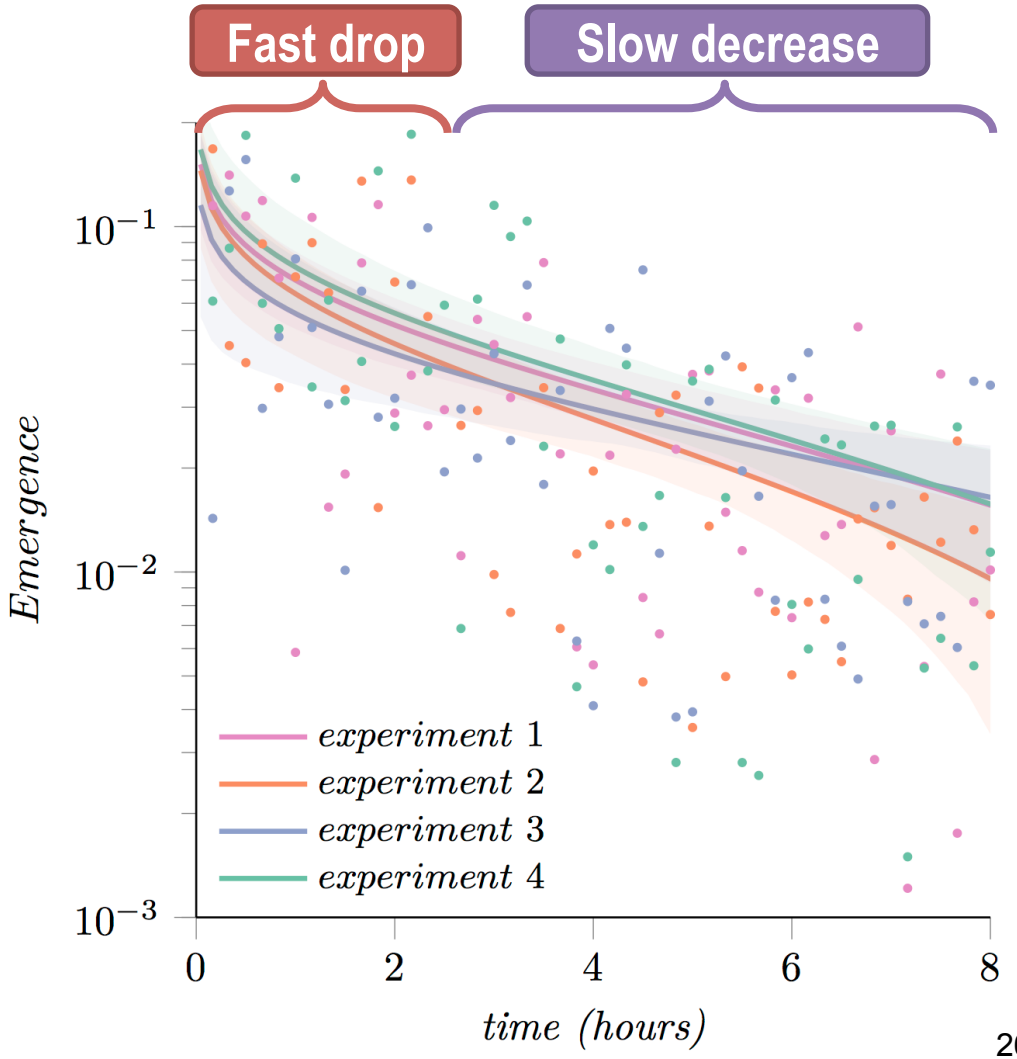
$$E = \int \frac{\partial D(q)}{\partial t} dq$$

- **Quantifies**
  - Direction of energy transfer
  - How much the distribution characterizing the system changes over time
  - Degree of emergence exhibited by microbial communities across multiple spatio-temporal scales



Emergence: Analysis

- **Emergence of aggregation spots represents the adaptation of Enterobacter cloacae to scarce food resources**
- **Emergence shows**
  - A fast drop at the beginning of the investigation period and
  - A slow decrease afterwards at longer times
- **Minimum energy principle**
  - If a system transitions to an equilibrium state while conserving the entropy, then the system minimizes its internal energy

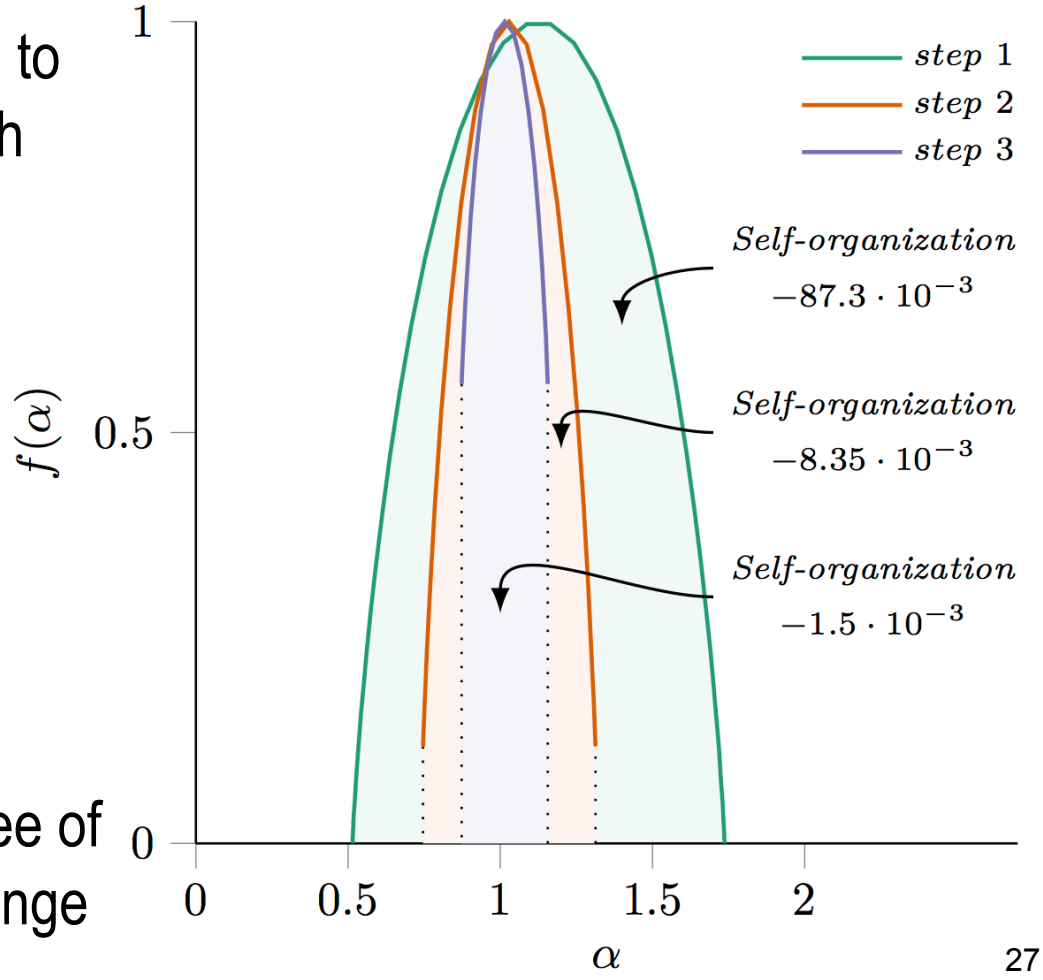


Self-organization: Definition

- **Self-organization quantifies**
  - How close a system is to the perfect order or self-similar distribution across all regions and observation scales
  - Ability of a swarm to converge to an ordered state solely through local interactions

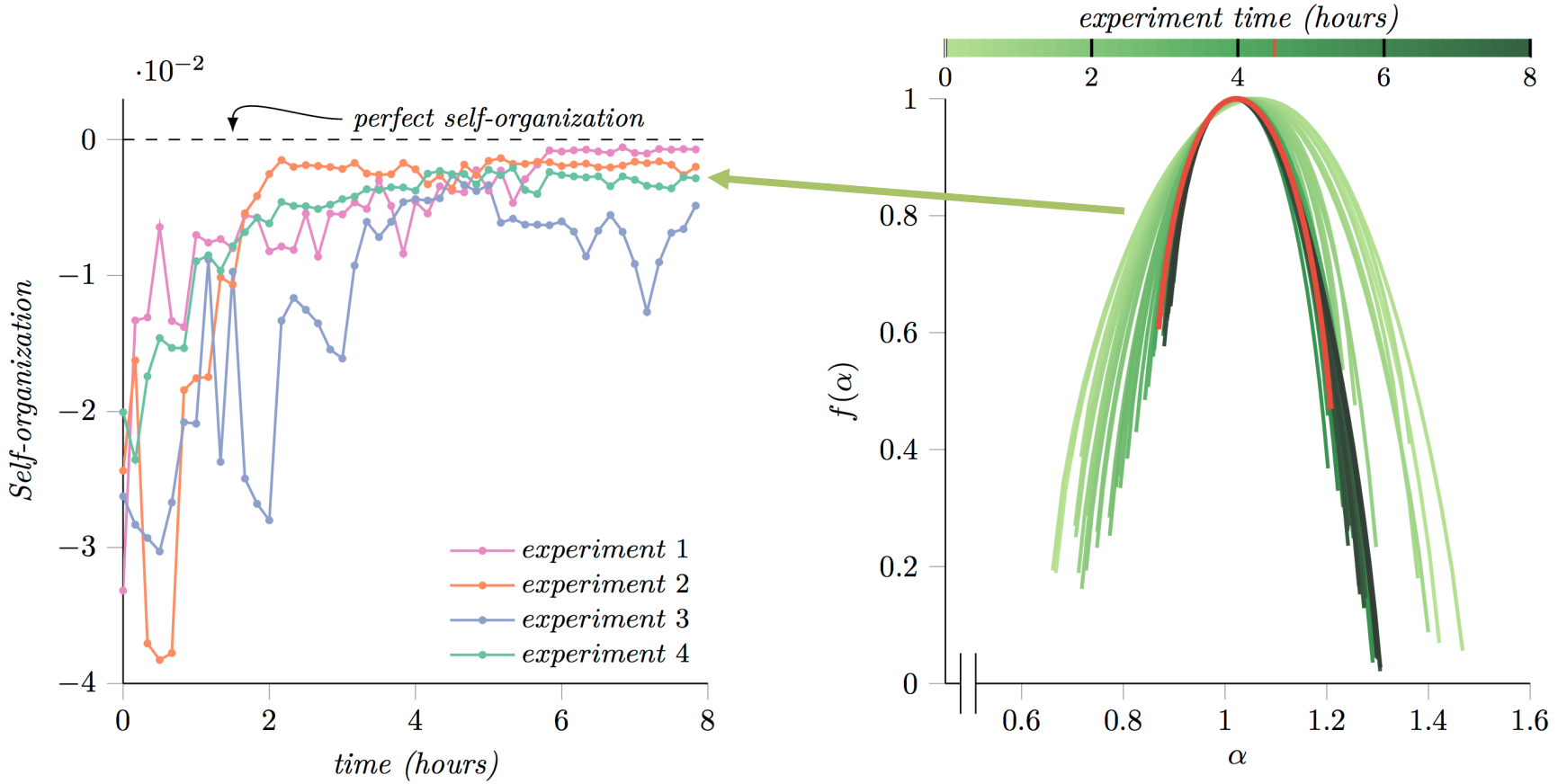
$$S = -\int (\alpha_0 - \alpha)^2 f(\alpha) d\alpha$$

  - Measures the deviation of multi-fractal spectrum from a peak α_0 (representing order)
  - Zero emergence implies degree of self-organization does not change



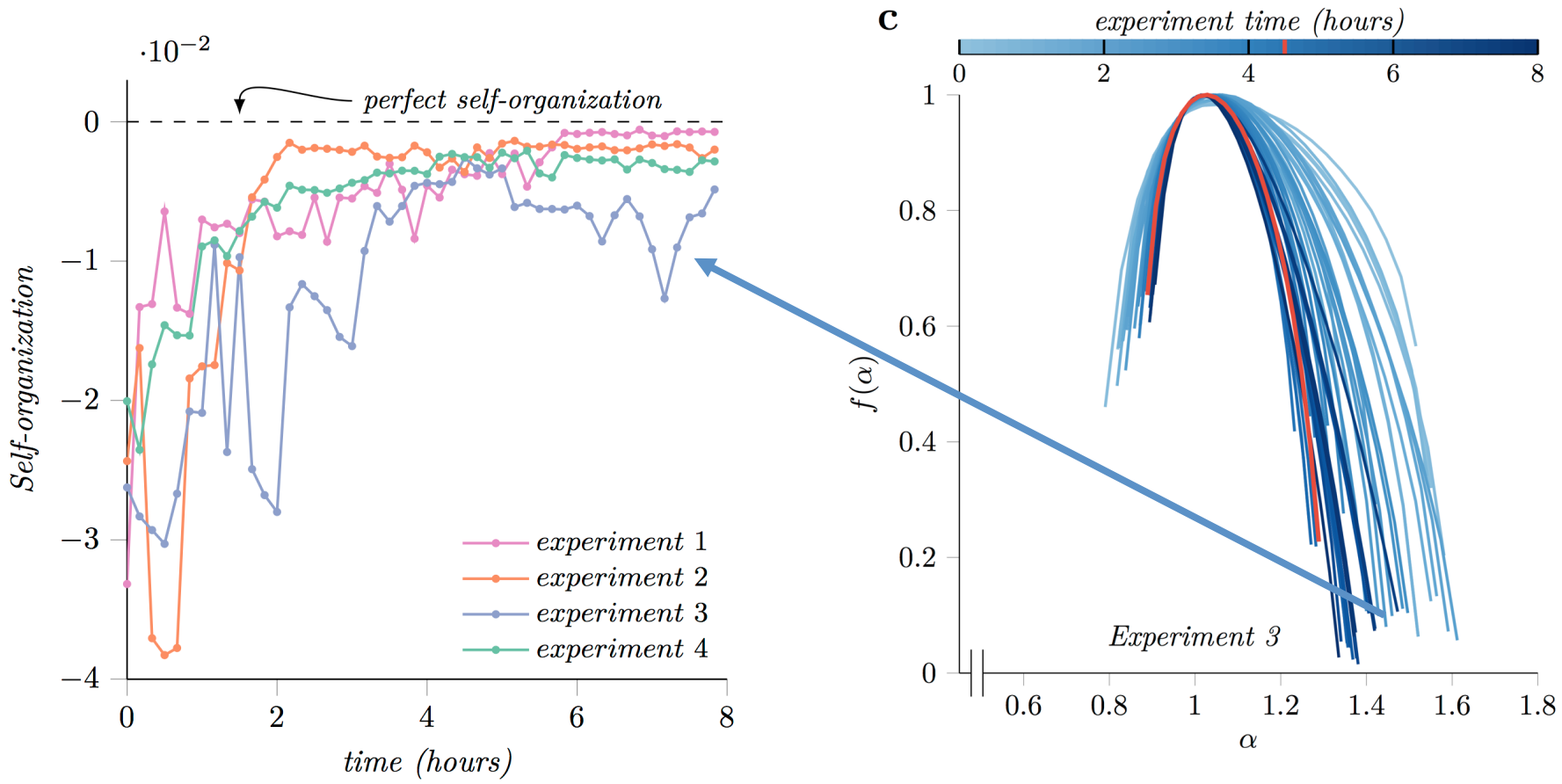
Self-Organization: Analysis (I)

- **Self-organization**
  - Increases towards the end of experiment
  - Suggests that spots formed later in time followed the rule impose by early spots and not formed at a random



Self-Organization: Analysis (II)

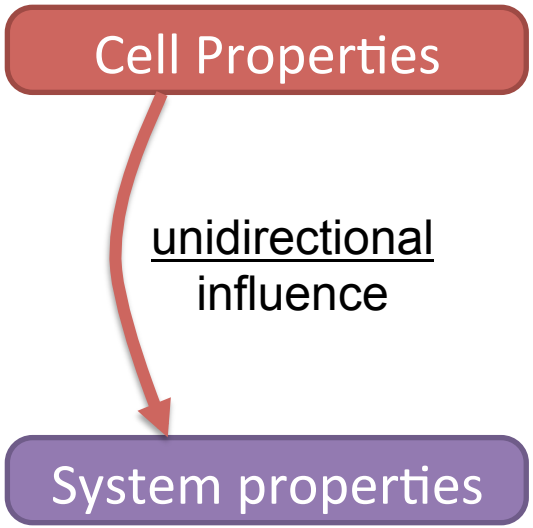
- **Degree of self-organization in experiment 3 fluctuates more since the right part of the spectrum exhibits higher variation compared to the other experiments**



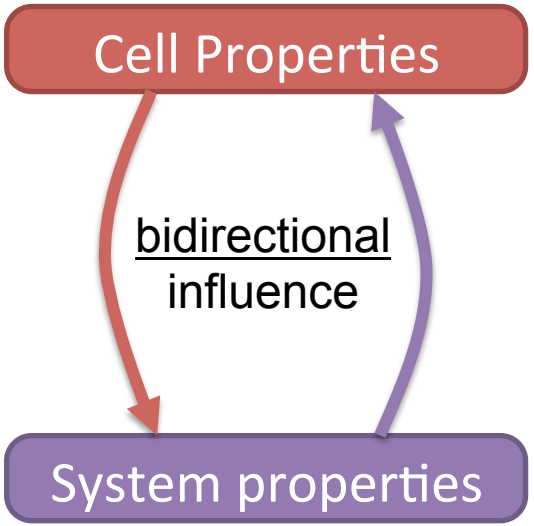
More on Emergence Analysis (I)

- **Causal emergence changes the dynamics complexity**

Non-causal emergence



Causal emergence



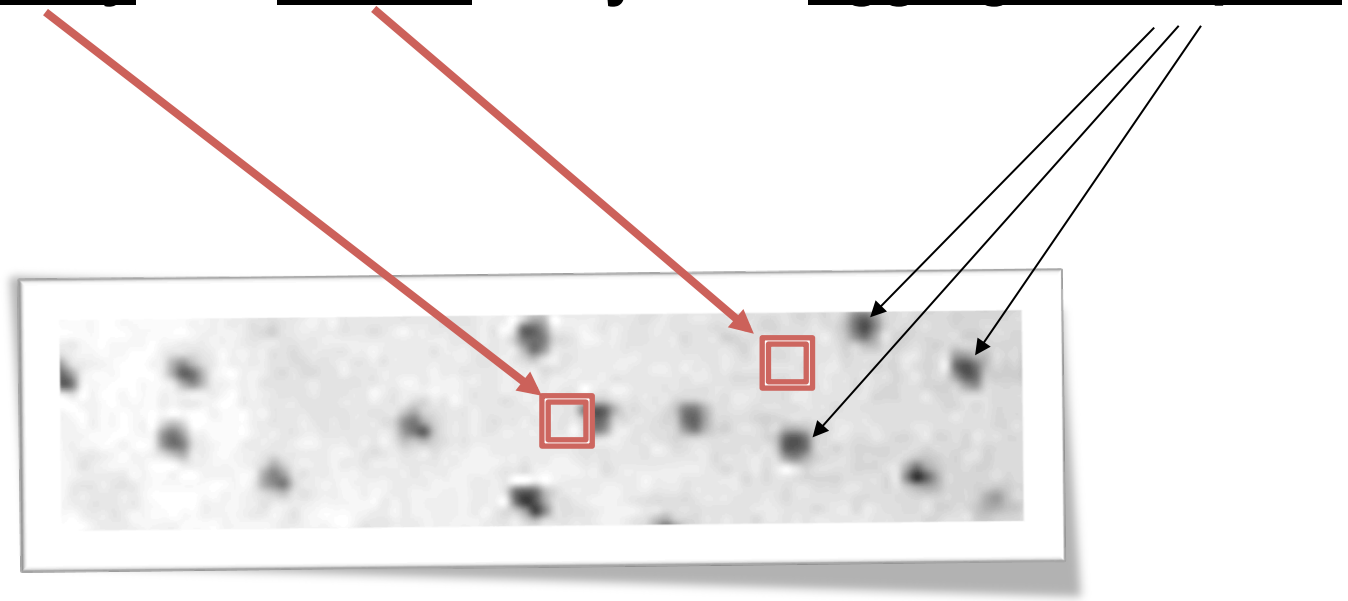
- Bidirectional influence represents a feedback loop that impacts the dynamics of the system by either
  - increasing the complexity (chaotic behavior), or
  - stabilizing the dynamics (restricted behavior)

More on Emergence Analysis (II)

- **Is bacteria aggregation a causal or a non-causal process?**

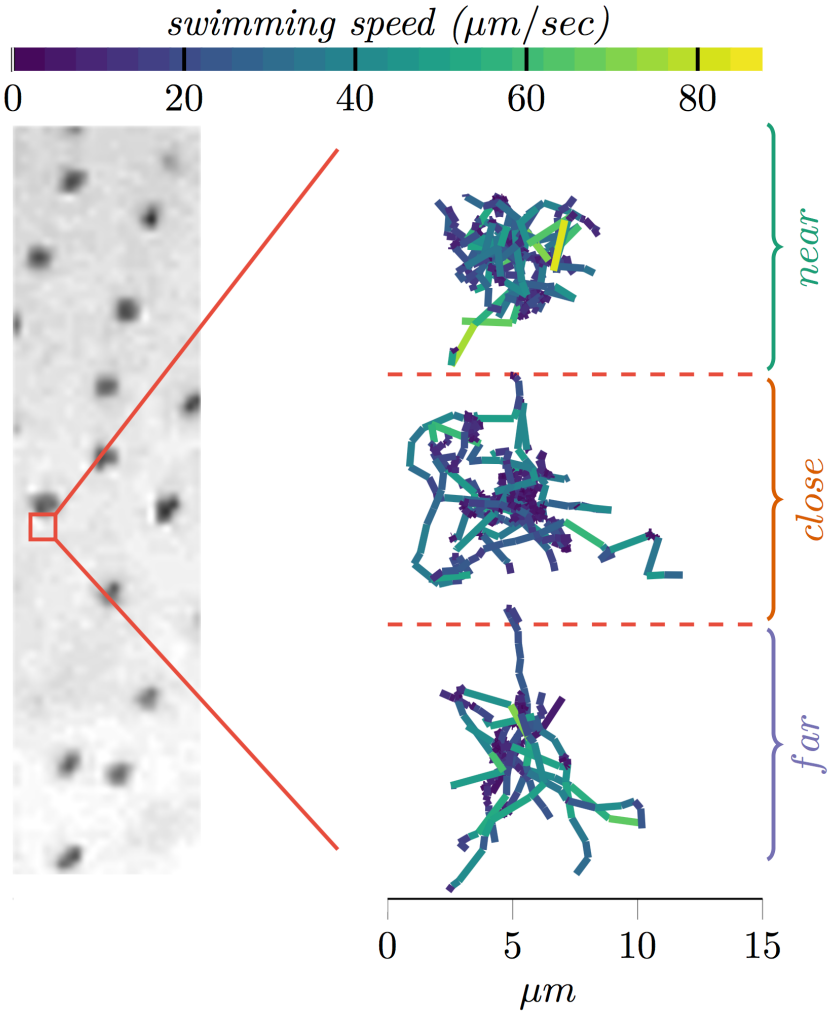
Investigate swimming behavior

in the vicinity and farther away from aggregation spots



Emergence Influences Dynamics

- **Near the spot the swimming trajectories become more concentrated around their initial position**

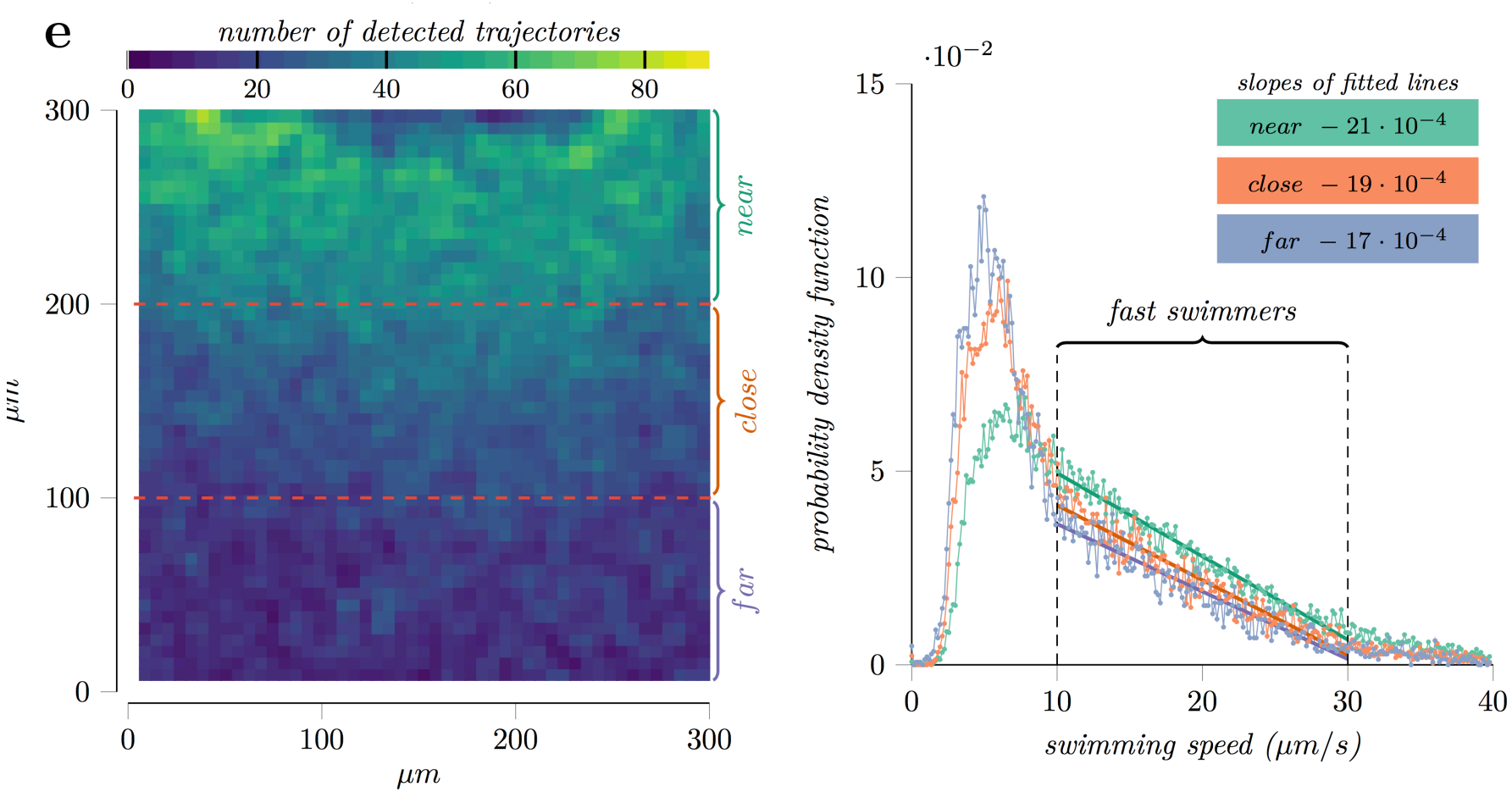


Shorter swimming runs

Longer swimming runs

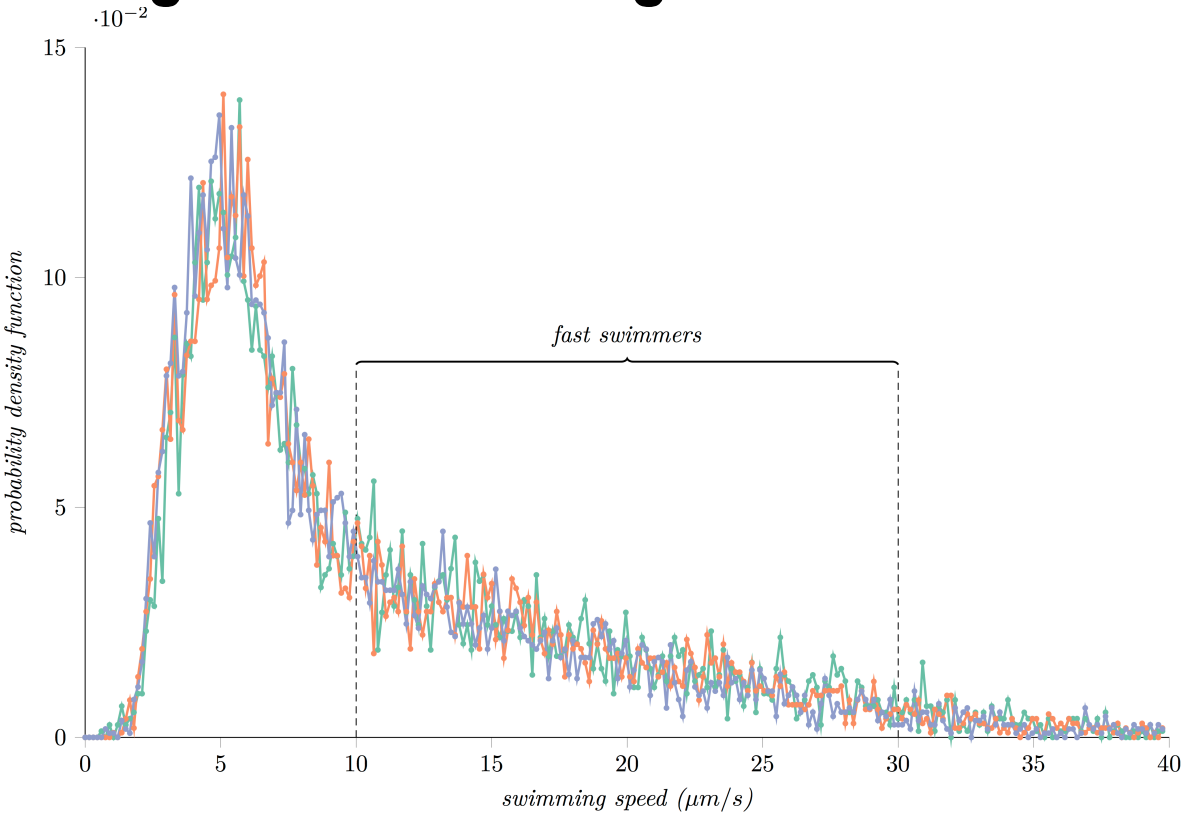
More on Emergence Analysis (III)

- **Bacteria cells swim faster in the vicinity of aggregation spot despite high cell concentration and high chance of collision**



More on Emergence Analysis (IV)

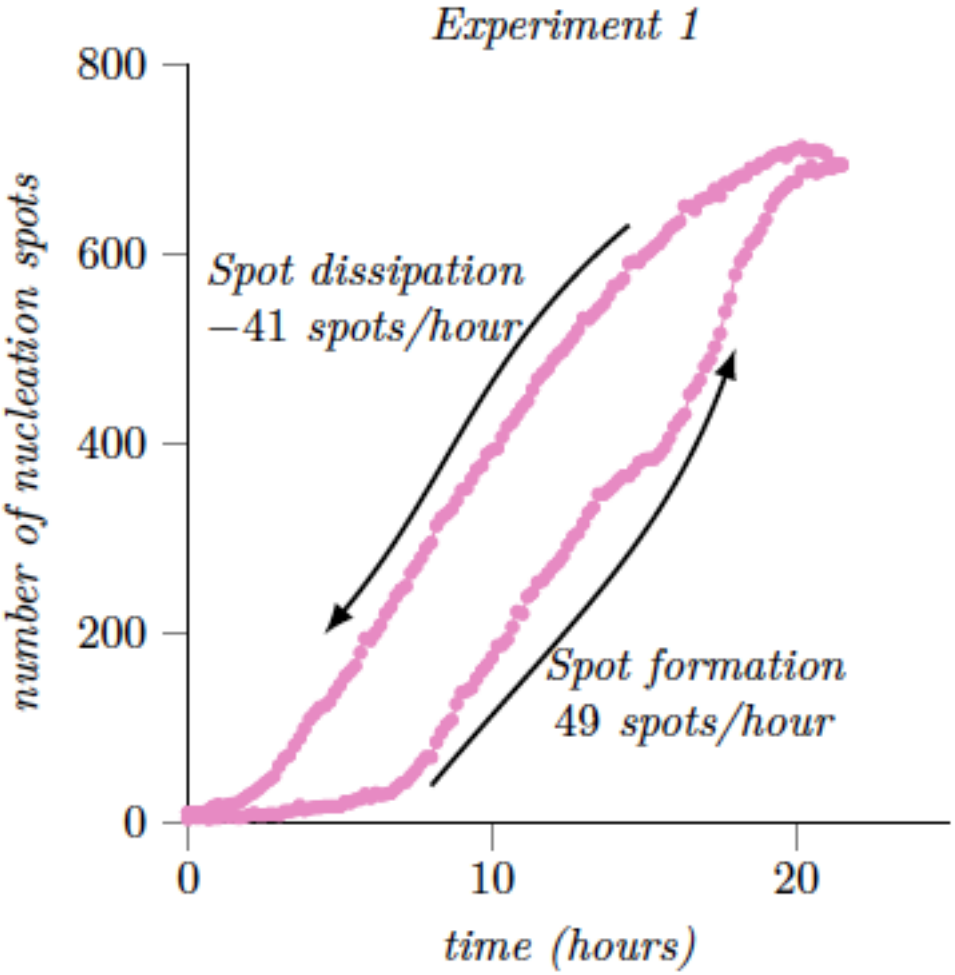
- **No difference in swimming behavior for regions between the aggregation spots**
- **Aggregation spots form as result of causal emergence**
  - Closer to aggregation spot the trajectories are more interrupted and concentrated around a point



  - Bacteria speed increases closer to the spot despite high cell density and chance of collision
  - No such behavior is observed for the regions in the middle between two neighboring spots

Hysteresis Behavior

- **Enterobacter cloacae communities exhibit a hysteresis behavior when transitioning between aggregate and non-aggregate states**
  - Hysteresis defines the dependence of a system state on its history (memory)
  - Spot formation
    - Transition from non-aggregate state to aggregate state has a speed of approximately 49 spots/h
  - Spot dissipation
    - Transition from aggregate to non-aggregate state has a speed of approximately 41 spots/h



Conclusions and Future Work

- **Current work**
  - Collect and record pattern formation in natural coliforms
  - Provide a non-Euclidean geometric characterization of patterns formed by natural isolates
  - Record genomic diversity and its impact on pattern formation
  - Test experimental hints through a computational model of microbial communities
  - Provide a statistical physics and information theory framework for quantifying computation and communication complexity
  - Quantify impact of single-cell heterogeneity on pattern formation
  - Quantify the robustness of pattern formation in mixed cell populations

Thank you!

More info at http://ceng.usc.edu/cps/